

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:47:29 ; Search time 6.90909 Seconds  
(without alignments)  
125.302 Million cell updates/sec

Title: US-09-761-636A-10  
Perfect score: 50  
Sequence: 1 CSVPLTSVC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	86.0	1487	2 S48719	phospholipase-A(2)
2	41	82.0	349	2 AD0990	probable membrane
3	41	82.0	364	2 C86015	hypothetical prote
4	41	82.0	376	2 S47693	hypothetical 38.5K
5	41	82.0	376	2 C91169	hypothetical prote
6	39	78.0	432	2 F64484	probable transamin
7	38	76.0	22	2 B55538	uvrC protein - pse
8	36	72.0	591	2 S54788	calcium-stimulated
9	36	72.0	2180	2 T29764	hypothetical prote
10	35	70.0	91	2 D97130	transcription regu
11	35	70.0	124	2 A37931	napin - Swedish tu
12	35	70.0	127	2 S65447	allergen Sin a I -
13	35	70.0	133	1 NWRP1	napin 1 precursor
14	35	70.0	145	2 S65478	allergen Sin a I (
15	35	70.0	145	2 S65481	allergen Sin a I (
16	35	70.0	145	2 S65482	allergen Sin a I (
17	35	70.0	145	2 S65480	hypothetical prote
18	35	70.0	152	2 T34649	napin B99 precursor
19	35	70.0	155	2 PS0426	napin 2 precursor
20	35	70.0	178	1 NWRP2	napin B - rape
21	35	70.0	178	2 S07828	napin precursor (n
22	35	70.0	178	2 A25997	2S storage protein
23	35	70.0	178	2 S25137	2S storage protein
24	35	70.0	178	2 S25130	2S storage protein
25	35	70.0	178	2 S25134	2S storage protein
26	35	70.0	180	2 S10018	napin (clone Bgna)
27	35	70.0	180	2 S52025	napin (clones BNMN
28	35	70.0	621	1 YRNC	monophenol monooxy
29	34	68.0	92	2 S35592	major allergen lar

30	34	68.0	141	2 T12116	NADH2 dehydrogenas
31	34	68.0	178	2 S15382	napin napB - rape
32	34	68.0	192	2 B96583	hypothetical prote
33	34	68.0	265	2 T11729	NADH2 dehydrogenas
34	34	68.0	270	2 T11687	NADH2 dehydrogenas
35	34	68.0	273	2 T11702	NADH2 dehydrogenas
36	34	68.0	277	2 T11697	NADH2 dehydrogenas
37	34	68.0	278	2 T11731	NADH2 dehydrogenas
38	34	68.0	376	2 S70841	hypothetical trans
39	34	68.0	377	2 S77634	exopolysaccharide
40	34	68.0	386	2 A57066	prostaglandin recep
41	34	68.0	418	2 E71716	proline/betaine tr
42	34	68.0	565	2 T08794	hypothetical prote
43	34	68.0	659	1 JC4365	arginine-tRNA liga
44	34	68.0	661	1 JN0870	arginine-tRNA liga
45	34	68.0	669	2 T17122	hypothetical prote

ALIGNMENTS

RESULT 1

S48719

phospholipase-A(2) receptor protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 20-Jun-2000

C:Accession: S48719

R:Higashino, K.; Ishizaki, J.; Kishino, J.; Ohara, O.; Arita, H.

Eur. J. Biochem. 225, 375-382, 1994

A:Title: Structural comparison of phospholipase-A(2)-binding regions in phospholipase-A

A:Reference number: S48719; MUID:95010128; PMID:7925459

A:Accession: S48719

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1487 <HIG>

A:Cross-references: GB:D30779; NID:gn375042; PIDN:BAA06443.1; PID:G691754

C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II r

F;181-222/Domain: fibronectin type II repeat homology <2F1>

F;380-503/Domain: C-type lectin homology <LCH>

Query Match 86.0%; Score 43; DB 2; Length 1487;  
Best Local Similarity 77.8%; Pred. No. 8;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVPLTSVC 9

Db 929 CSVPLPSIC 937

RESULT 2

AD0990

probable membrane protein STY4229 [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A>Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AD0990

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AD0990

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-349 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08048.1; PID:G16505028; GSPDB:GN00176

C:Genetics:

A:Gene: STY4229

C:Superfamily: conserved hypothetical protein HI0338

Query Match 82.0%; Score 41; DB 2; Length 349;

Best Local Similarity 100.0%; Pred. No. 4.9; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVPLTSVC 9  
DB 314 SVPLTSVC 321  
|||||

RESULT 3  
C86015  
hypothetical protein yhhT [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: C86015  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C86015  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-364 <STO>  
A:Cross-references: GB:AE005174; NID:g12518133; PIDN:AA058583.1; GSPDB:GN00145; UWGP:Z48  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: yhhT  
C:Superfamily: conserved hypothetical protein HI0338

Query Match 82.0%; Score 41; DB 2; Length 364;  
Best Local Similarity 100.0%; Pred. No. 5.1; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVPLTSVC 9  
DB 329 SVPLTSVC 336  
|||||

RESULT 4  
S47693  
hypothetical 38.5K protein (ftsY-nika intergenic region) - Escherichia coli (strain K-12)  
N:Alternate names: hypothetical protein o376  
C:Species: Escherichia coli  
C>Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 01-Mar-2002  
C:Accession: S47693; E65144  
R:Plunkett, G.  
submitted to the EMBL Data Library, March 1994  
A:Reference number: S47666  
A:Accession: S47693  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <PLU>  
A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AA018449.1; PID:g912460  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: E65144  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <BLAT>  
A:Cross-references: GB:AE000423; GB:U00096; NID:g1789880; PIDN:AA076499.1; PID:g1789885;  
A:Experimental source: strain K-12, substrain MGL655  
C:Genetics:  
A:Gene: yhhT  
A:Start codon: GTG  
C:Superfamily: conserved hypothetical protein HI0338

Query Match 82.0%; Score 41; DB 2; Length 376;  
Best Local Similarity 100.0%; Pred. No. 5.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVPLTSVC 9  
DB 341 SVPLTSVC 348  
|||||

RESULT 5  
C91169  
hypothetical protein EC94323 [imported] - Escherichia coli (strain O157:H7, substrain R  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C91169  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91169  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA037746.1; PID:g13363797; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 050952  
C:Genetics:  
A:Gene: EC94323  
C:Superfamily: conserved hypothetical protein HI0338

Query Match 82.0%; Score 41; DB 2; Length 376;  
Best Local Similarity 100.0%; Pred. No. 5.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVPLTSVC 9  
DB 341 SVPLTSVC 348  
|||||

RESULT 6  
F64484  
probable transaminase (EC 2.6.1.-) MJ1479 [similarity] - Methanococcus jannaschii  
N:Alternate names: alanine aminotransferase 2  
C:Species: Methanococcus jannaschii  
C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jun-2000  
C:Accession: F64484  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A:Reference number: A64300; MUID:96337999; PMID:8688087  
A:Accession: F64484  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-432 <BUL>  
A:Cross-references: GB:U67588; GB:L77117; NID:g1592111; PID:g1592118; TIGR:MJ1479; PID  
C:Genetics:  
A:Map position: FOR1449710-1451008  
C:Superfamily: aspartate transaminase  
C:Keywords: aminotransferase

Query Match 78.0%; Score 39; DB 2; Length 432;  
Best Local Similarity 77.8%; Pred. No. 14; Mismatches 2; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
DB 389 CSVPLTSVC 397  
|||||

RESULT 7  
B5538  
uvrC protein - Pseudomonas syringae pv. syringae (strain B728a) (fragment)  
C:Species: Pseudomonas syringae pv. syringae  
C>Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 08-Sep-1995

C;Accession: B55538  
 R;Rich, J.J.; Kinscherf, T.G.; Kitten, T.; Willis, D.K.  
 J. Bacteriol. 176, 7468-7475, 1994  
 A;Title: Genetic evidence that the gacA gene encodes the cognate response regulator for  
 A;Reference number: A55538; MUID:95095914; PMID:8002569  
 A;Accession: B55538  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-22 <RIC>  
 A;Cross-references: GB:U09767  
 C;Genetics:  
 A;Gene: uvrC

Query Match 76.0%; Score 38; DB 2; Length 22;  
 Best Local Similarity 66.7%; Pred. No. 1.4;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPASTVC 9  
 Db 13 CSVPASTVC 21  
 |||| :|||

RESULT 8  
 S54788  
 calcium-stimulated protein kinase - Chlamydomonas eugametos  
 C;Species: Chlamydomonas eugametos  
 C;Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 18-Jun-1999  
 C;Accession: S54788  
 R;Siederius, M.H.  
 submitted to the EMBL Data Library, May 1995  
 A;Description: Cloning a cDPK from Chlamydomonas eugametos.  
 A;Reference number: S54788  
 A;Accession: S54788  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-591 <SID>  
 A;Cross-references: EMBL:Z49233; NID:g806541; PIDN:CRA89202.1; PID:g806542  
 C;Superfamily: calcium-dependent protein kinase; calmodulin repeat homology; protein kin  
 C;Keywords: ATP; calcium binding; EF hand; serine/threonine-specific protein kinase; tan  
 F;150-410/Domain: protein kinase homology <KIN>  
 F;158-166/Region: protein kinase ATP-binding motif  
 F;453-485/Domain: calmodulin repeat homology <EF1>  
 F;489-521/Domain: calmodulin repeat homology <EF2>  
 F;525-557/Domain: calmodulin repeat homology <EF3>  
 F;559-591/Domain: calmodulin repeat homology <EF4>

Query Match 72.0%; Score 36; DB 2; Length 591;  
 Best Local Similarity 55.6%; Pred. No. 66;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CSVPPLTVC 9  
 Db 88 CKIPLTRAC 96  
 ||||| :|||

RESULT 9  
 T29764  
 hypothetical protein T21E3.3 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Aug-2000  
 C;Accession: T29764  
 R;Du, Z.; Le, T.T.  
 submitted to the EMBL Data Library, May 1997  
 A;Description: The sequence of C. elegans cosmid T21E3.  
 A;Reference number: Z20681  
 A;Accession: T29764  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-2180 <DUZ>  
 A;Cross-references: EMBL:AF003133; PIDN:AAB54138.1; GSPDB:GN00019; CESP:T21E3.3  
 A;Experimental source: strain Bristol N2; clone T21E3  
 C;Genetics:  
 A;Gene: CESP:T21E3.3

A;Map position: 1  
 C;Introns: 400/2; 1608/3; 1644/1; 1674/3; 1686/2; 1731/2; 2073/1; 2098/2; 2146/3  
 C;Superfamily: LDL receptor ligand-binding repeat homology  
 F;15-49/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F;51-85/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F;90-132/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F;149-185/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F;190-225/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
 F;229-272/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
 F;285-316/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
 F;323-359/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
 F;369-405/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
 F;816-856/Domain: LDL receptor ligand-binding repeat homology <LDL10>  
 F;861-903/Domain: LDL receptor ligand-binding repeat homology <LDL11>  
 F;908-943/Domain: LDL receptor ligand-binding repeat homology <LDL12>  
 F;948-987/Domain: LDL receptor ligand-binding repeat homology <LDL13>  
 F;993-1027/Domain: LDL receptor ligand-binding repeat homology <LDL14>  
 F;1031-1063/Domain: LDL receptor ligand-binding repeat homology <LDL15>  
 F;1126-1165/Domain: LDL receptor ligand-binding repeat homology <LDL16>  
 F;1172-1209/Domain: LDL receptor ligand-binding repeat homology <LDL17>  
 F;1213-1246/Domain: LDL receptor ligand-binding repeat homology <LDL18>

Query Match 72.0%; Score 36; DB 2; Length 2180;  
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 VPLTVC 9  
 Db 922 IPLTVC 928  
 :|||

## RESULT 10

D97130  
 transcription regulator (phage-related) (Xre family) [imported] - Clostridium acetobutylicum  
 C;Species: Clostridium acetobutylicum  
 C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
 C;Accession: D97130  
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C  
 A;Reference number: A96900; MUID:21359325; PMID:21359325  
 A;Accession: D97130  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-91 <KUR>  
 A;Cross-references: GB:AE001437; PIDN:AAK79831.1; PID:gi5024845; GSPDB:GN00168  
 A;Experimental source: Clostridium acetobutylicum ATCC824  
 C;Genetics:  
 A;Gene: CAC1867

Query Match 70.0%; Score 35; DB 2; Length 91;  
 Best Local Similarity 66.7%; Pred. No. 18;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPPLTVC 9  
 Db 54 CSIGYTSVC 62  
 ||| :|||

## RESULT 11

A37931  
 napin - Swedish turnip (fragments)  
 C;Species: Brassica napus var. rapifera (Swedish turnip, rutabaga)  
 C;Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 12-Apr-1995  
 C;Accession: A37931  
 R;Svendsen, I.; Nicolova, D.; Goshev, I.; Genov, N.  
 Carlsberg Res. Commun. 54, 231-239, 1989  
 A;Title: Isolation and characterization of a trypsin inhibitor from the seeds of kohlrab  
 A;Reference number: A37931; MUID:91282906; PMID:2490369  
 A;Accession: A37931  
 A;Status: preliminary  
 A;Molecule type: protein

A;Residues: 1-124 <SVE>

C;Superfamily: wheat alpha-amylase inhibitor

Query Match 70.0%; Score 35; DB 2; Length 124;  
Best Local Similarity 55.6%; Pred. No. 24;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPPLTSVC 9

Db 106 CNIPQVSVC 114

RESULT 12

S65447

allergen Sin a I - white mustard

N;Alternate names: allergen Sin a I large chain

C;Species: Sinapis alba (white mustard)

C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-1999

C;Accession: S65447; S01792

R;Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.

Eur. J. Biochem. 237, 827-832, 1996

A;Title: Expression in Escherichia coli of Sin a 1, the major allergen from mustard.

A;Reference number: S65447; MUID:96235251; PMID:8647131

A;Accession: S65447

A;Molecule type: protein

A;Residues: 1-127 <GON>

A;Experimental source: seed

R;Menendez-Arias, L.; Monco, I.; Dominguez, J.; Rodriguez, R.

Eur. J. Biochem. 177, 159-166, 1988

A;Title: Primary structure of the major allergen of yellow mustard (Sinapis alba L.) seed

A;Reference number: S01791; MUID:89030681; PMID:3181153

A;Accession: S01792

A;Molecule type: protein

A;Residues: 40-127 <MEN>

C;Superfamily: wheat alpha-amylase inhibitor

C;Keywords: seed; storage protein

Query Match

Best Local Similarity 70.0%; Score 35; DB 2; Length 127;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPPLTSVC 9

Db 109 CNIPQVSVC 117

RESULT 13

NWRP1

napin 1 precursor - rape (fragment)

N;Alternate names: 1-7S seed storage protein

C;Species: Brassica napus (rape)

C;Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 18-Jun-1999

C;Accession: A01330

R;Crouch, M.L.; Tenbarge, K.M.; Simon, A.E.; Ferl, R.

J. Mol. Appl. Genet. 2, 273-283, 1983

A;Title: cDNA clones for Brassica napus seed storage proteins: evidence from nucleotide

A;Reference number: A92836; MUID:84113267; PMID:6683334

A;Accession: A01330

A;Molecule type: mRNA

A;Residues: 1-133 <CRO>

A;Cross-references: GB:K01544; NID:G167174; PIDN:AAA33005.1; PID:G167175

A;Experimental source: cv. Tower

C;Superfamily: wheat alpha-amylase inhibitor

C;Keywords: seed; storage protein

F;1-30/Product: napin 1 small chain (fragment) #status predicted <SCH>

F;50-130/Product: napin 1 large chain #status predicted <LCH>

Query Match

Best Local Similarity 70.0%; Score 35; DB 1; Length 133;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPPLTSVC 9

Db 109 CNIPQVSVC 117

Db 114 CNIPQVSVC 122

RESULT 14

S65478

allergen Sin a I (clone SIN5) - white mustard (fragment)

C;Species: Sinapis alba (white mustard)

C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 16-Jul-1999

C;Accession: S65478

R;Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.

Eur. J. Biochem. 237, 827-832, 1996

A;Title: Expression in Escherichia coli of Sin a 1, the major allergen from mustard.

A;Reference number: S65447; MUID:96235251; PMID:8647131

A;Accession: S65478

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-145 <GON>

A;Cross-references: EMBL:X91798; NID:G1009441; PIDN:CAA62908.1; PID:G1009442

A;Experimental source: seed

C;Genetics:

A;Gene: sin5

C;Superfamily: wheat alpha-amylase inhibitor

C;Keywords: seed; storage protein

Query Match

Best Local Similarity 70.0%; Score 35; DB 2; Length 145;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPPLTSVC 9

Db 127 CNIPQVSVC 135

RESULT 15

S65481

allergen Sin a I (clone SIN3) - white mustard (fragment)

C;Species: Sinapis alba (white mustard)

C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 16-Jul-1999

C;Accession: S65481

R;Gonzalez de la Pena, M.A.; Monsalve, R.I.; Batanero, E.; Villalba, M.; Rodriguez, R.

Eur. J. Biochem. 237, 827-832, 1996

A;Title: Expression in Escherichia coli of sin a 1, the major allergen from mustard.

A;Reference number: S65447; MUID:96235251; PMID:8647131

A;Accession: S65481

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-145 <GON>

A;Cross-references: EMBL:X91801; NID:G1009437; PIDN:CAA62911.1; PID:G1009438

A;Experimental source: seed

C;Genetics:

A;Gene: sin3

C;Superfamily: wheat alpha-amylase inhibitor

C;Keywords: seed; storage protein

Query Match

Best Local Similarity 70.0%; Score 35; DB 2; Length 145;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPPLTSVC 9

Db 127 CNIPQVSVC 135

Search completed: September 5, 2004, 10:01:20

Job time : 7.90909 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:38:39 ; Search time 4.09091 Seconds  
(without alignments)  
114.554 Million cell updates/sec

Title: US-09-761-636A-10  
Perfect score: 50  
Sequence: 1 CSVPLTSVC 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	41	82.0	349	1	YHHT_ECOLI
2	38	76.0	150	1	UVRC_PSESY
3	35	70.0	127	1	ALL1_SINAL
4	35	70.0	133	1	2SS1_BRANA
5	35	70.0	178	1	2SS2_BRANA
6	35	70.0	180	1	2SS3_BRANA
7	35	70.0	216	1	1KGD_MACMU
8	35	70.0	620	1	TYRO_NEUCR
9	35	70.0	660	1	TYRO_MOUSE
10	34	68.0	129	1	ALL1_BRAJU
11	34	68.0	178	1	2SSB_BRANA
12	34	68.0	377	1	EP42_RALSO
13	34	68.0	386	1	PI2R_HUMAN
14	34	68.0	574	1	CIK3_HUMAN
15	34	68.0	586	1	CSPI_RAT
16	34	68.0	660	1	SYR_HUMAN
17	34	68.0	661	1	SYR_CRILLO
18	33	66.0	100	1	GATC_RICCN
19	33	66.0	311	1	HEM3_NEIMA
20	33	66.0	311	1	HEM3_NEIMB
21	33	66.0	572	1	NUSM_LOCM1
22	33	66.0	1458	1	PA2R_RABIT
23	32	64.0	178	1	YGYS_YEAST
24	32	64.0	216	1	1KGD_HUMAN
25	32	64.0	308	1	SAL_SILAS
26	32	64.0	318	1	DEOC_MOUSE
27	32	64.0	326	1	VEGD_RAT
28	32	64.0	353	1	RN28_HUMAN
29	32	64.0	354	1	DDL_CANEF
30	32	64.0	354	1	VEGD_HUMAN
31	32	64.0	358	1	VEGD_MOUSE
32	32	64.0	378	1	ADHQ_RABIT
33	32	64.0	391	1	ADHF_HUMAN

34 32 64.0 397 1 NXF5\_HUMAN  
35 32 64.0 403 1 G3PA\_MAIZE  
36 32 64.0 427 1 GGNT\_BOVIN  
37 32 64.0 436 1 RN29\_HUMAN  
38 32 64.0 451 1 G3PB\_PEA  
39 32 64.0 493 1 ALG8\_PSES  
40 32 64.0 502 1 PYRG\_MYCCA  
41 32 64.0 598 1 NUSM\_BRALA  
42 32 64.0 598 1 Y427\_HUMAN  
43 32 64.0 599 1 NUSM\_BRAFL  
44 32 64.0 616 1 NXF1\_COTJA  
45 32 64.0 660 1 VNCS\_PAVEN

Q9hlb4 homo sapien  
P09315 zea mays (m  
Q92180 bos taurus  
Q9byv6 homo sapien  
P12859 pisum sativ  
Q887p9 pseudomonas  
Q48965 mycoplasma  
O79422 branchiosto  
O43310 homo sapien  
O47430 branchiosto  
P58797 coturnix co  
P18547 porcine par

## ALIGNMENTS

RESULT 1  
YHHT\_ECOLI STANDARD; PRT; 349 AA.  
AC P37622; P76700; O8X6P3;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein yhhf.  
GN YHHT OR B3474 OR Z4848 OR ECS4323.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=94316500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
RT region from 76.0 to 81.5 minutes.";  
RL Nucleic Acids Res. 22:2576-2586(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RL Nature 409:529-533(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
Han C.-G., Ohsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22(2001).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY. STRONG, TO  
CC H.INFLUENZAE HI0338.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC ENBL; U00039; AAB18449.1; ALT INIT.
DR EMBL; AE000423; AAC76499.1; ALT INIT.
DR ENBL; AE005570; AAG58583.1; ALT INIT.
DR ENBL; AP002565; BAB37746.1; ALT INIT.
DR EcoGene; EG12220; Yhht.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
SQ SEQUENCE 349 AA; 38522 MW; 275ADBADA4FDS8257 CRC64;

Query Match 82.0%; Score 41; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 0.95; 0; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 0;

Qy 2 SVPLTSVC 9
Db 314 SVPLTSVC 321

RESULT 2
UVR_C PSSY STANDARD; PRT; 150 AA.
AC Q52377;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UVRABC system protein C (UvrC protein) (Excinuclease ABC subunit C)
DE (Fragment).
GN UVR_C.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B728a;
RX MEDLINE=95095914; PubMed=8002569;
RA Rich J.J., Kinscherf T.G., Kitten T., Willis D.K.;
RT "Genetic evidence that the gacA gene encodes the cognate response
regulator for the lens sensor in Pseudomonas syringae.";
RL J. Bacteriol. 176:7468-7475(1994).
CC -!- FUNCTION: The UVRABC repair system catalyzes the recognition and
processing of DNA lesions. UvrC both incises the 5' and 3' sides
of the lesion. The N-terminal half is responsible for the 3'
incision and the C-terminal half is responsible for the 5'
incision (By similarity).
CC -!- SUBUNIT: Interacts with uvrB in an incision complex (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the uvrC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR ENBL; U09767; AAA65232.1; -
DR PIR; B55538; B55538.
DR HAMAP; MF 00203; -; 1.
DR InterPro; IPR000305; UvrC_N.
DR Pfam; PF01541; Exci_endo_N; 1.
```

```
DR SMART; SM00465; GIYC; 1.
DR PROSITE; PS0164; UVR_C 1; 1.
KW SOS response; Excision nuclease; DNA repair; DNA recombination;
KW DNA excision.
FT NON_TER 150 150
SQ SEQUENCE 150 AA; 16999 MW; ACC38718B5F1E2B1 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 150;
Best Local Similarity 66.7%; Pred. No. 1.5; 1; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 2;

Qy 1 CSVPLTSVC 9
Db 13 CSVPLTSVC 21

RESULT 3
ALLI_SINAL STANDARD; PRT; 127 AA.
AC P15322;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Allergen Sin a 1, small and large chains (Sin a 1).
OS Sinapis alba (White mustard) (Brassica hirta).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Sinapis.
OX NCBI_TaxID=3728;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=89030681; PubMed=3181153;
RA Menendez-Arias L., Moneo I., Dominguez J., Rodriguez R.;
RT "Primary structure of the major allergen of yellow mustard (Sinapis
alba L.) seed, Sin a 1.";
RL Eur. J. Biochem. 177:159-166(1988).
CC -!- FUNCTION: This is a 2S seed storage protein.
CC -!- SUBUNIT: The protein consists of two chains linked by disulfide
bonds.
CC -!- ALLERGEN: Causes an allergic reaction in human. Causes cabbage
allergy.
CC -!- SIMILARITY: Belongs to the 2S seed storage albumins family.
CC PIR; S01791; S01791.
CC PIR; S65447; S65447.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR000617; Napin.
DR PRINTS; PRO0496; NAPIN.
DR ProDom; PD002498; Napin; 1.
DR SMART; SM00499; AAI; 1.
DR Allergen; Seed storage protein.
KW CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 127 LARGE CHAIN.
FT VARIANT 6 6 R -> G.
SQ SEQUENCE 127 AA; 14180 MW; 4CD920284F04BEFD CRC64;

Query Match 70.0%; Score 35; DB 1; Length 127;
Best Local Similarity 55.6%; Pred. No. 5; 2; Indels 0; Gaps 0;
Matches 5; Conservative 2; Mismatches 2;

Qy 1 CSVPLTSVC 9
Db 109 CNIPQSVVC 117

RESULT 4
2SS1_BRANA STANDARD; PRT; 133 AA.
ID 2SS1_BRANA
AC P01091;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

DE Napin 1 precursor (1.7S seed storage protein) (Fragment).  
 OS Brassica napus (Rape).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Tower;  
 RX MEDLINE=84113267; PubMed=6689334;  
 RA Crouch M.L., Tenberge K.M., Simon A.E., Ferl R.;  
 RT "cDNA clones for Brassica napus seed storage proteins: evidence from  
 RT nucleotide sequence analysis that both subunits of napin are cleaved  
 RT from a precursor polypeptide.";  
 RL J. Mol. Appl. Genet. 2:273-283(1983).  
 CC -!- FUNCTION: The small, basic, water-soluble napins are one of the  
 CC two major kinds of storage proteins synthesized in the seed during  
 CC its maturation.  
 CC -!- SUBUNIT: The mature protein consists of a small and a large chain  
 CC linked by disulfide bonds.  
 CC -!- TISSUE SPECIFICITY: Cotyledons and the axis.  
 CC -!- SIMILARITY: Belongs to the 2S seed storage albumins family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K01544; AAA33005.1; -;  
 DR PIR; A01330; NWRP1.  
 DR InterPro; IPR003612; AAI.  
 DR Pfam; PF00234; tryp\_alpha\_aml; 1.  
 DR PRINTS; PR00496; NAFIN.  
 DR ProDom; PD002498; Napin; 1.  
 DR SMART; SM00499; AAI; 1.  
 DR Seed storage protein; Multigene family.  
 FT NON TER 1 1  
 FT CHAIN <1 30 SMALL CHAIN.  
 FT PROPEP 31 49  
 FT CHAIN 50 130 LARGE CHAIN.  
 SQ SEQUENCE 133 AA; 15294 MW; 5FDFC7ECE3E22ACB CRC64;  
 Query Match 70.0%; Score 35; DB 1; Length 133;  
 Best Local Similarity 55.6%; Pred. No. 5.3;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CSVPPLTSVC 9  
 Db 114 CNIPQVSVC 122  
 RESULT 5  
 ID 2SS2\_BRANA STANDARD; PRT; 178 AA.  
 AC P01090;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE Napin 2 precursor (1.7S seed storage protein).  
 OS Brassica napus (Rape).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87308224; PubMed=3624251;  
 RA Josefsson L.-G., Lenman M., Ericson M.L., Raak L.;  
 RT "Structure of a gene encoding the 1.7 S storage protein, napin, from

RT Brassica napus.";  
 RL J. Biol. Chem. 262:12196-12201(1987).  
 RN [2]  
 RP REVISIONS.  
 RA Josefsson L.-G.;  
 RL Submitted (JUL-1987) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87033665; PubMed=3771543;  
 RA Ericson M.L., Roedin J., Lenman M., Glimelius K., Josefsson L.-G.,  
 RA Raak L.;  
 RT "Structure of the rapeseed 1.7 S storage protein, napin, and its  
 RT precursor.";  
 RL J. Biol. Chem. 261:14576-14581(1986).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Tower;  
 RX MEDLINE=84113267; PubMed=6689334;  
 RA Crouch M.L., Tenberge K.M., Simon A.E., Ferl R.;  
 RT "cDNA clones for Brassica napus seed storage proteins: evidence from  
 RT nucleotide sequence analysis that both subunits of napin are cleaved  
 RT from a precursor polypeptide.";  
 RL J. Mol. Appl. Genet. 2:273-283(1983).  
 CC -!- FUNCTION: The small, basic, water-soluble napins are one of the  
 CC two major kinds of storage proteins synthesized in the seed during  
 CC its maturation.  
 CC -!- SUBUNIT: The mature protein consists of a small and a large chain  
 CC linked by disulfide bonds.  
 CC -!- TISSUE SPECIFICITY: Cotyledons and the axis.  
 CC -!- SIMILARITY: Belongs to the 2S seed storage albumins family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; K01545; AAA33006.1; -;  
 DR EMBL; J02586; AAA32997.1; -;  
 DR EMBL; J02798; AAA87348.1; -;  
 DR PIR; A01329; NWRP2.  
 DR PIR; A29801; A25997.  
 DR InterPro; IPR003612; AAI.  
 DR InterPro; IPR000617; Napin.  
 DR Pfam; PF00234; tryp\_alpha\_aml; 1.  
 DR PRINTS; PR00496; NAFIN.  
 DR ProDom; PD002498; Napin; 1.  
 DR SMART; SM00499; AAI; 1.  
 DR Seed storage protein; Signal; Multigene family.  
 FT SIGNAL 1 21  
 FT PROPEP 22 38  
 FT CHAIN 39 74 SMALL CHAIN.  
 FT PROPEP 75 94  
 FT CHAIN 95 175 LARGE CHAIN.  
 FT CONFLICT 37 37 D -> N (IN REF. 4).  
 FT CONFLICT 76 76 S -> N (IN REF. 4).  
 SQ SEQUENCE 178 AA; 20104 MW; 734E561971B539FF CRC64;  
 Query Match 70.0%; Score 35; DB 1; Length 178;  
 Best Local Similarity 55.6%; Pred. No. 7.2;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CSVPPLTSVC 9  
 Db 159 CNIPQVSVC 167  
 RESULT 6  
 ID 2SS3\_BRANA STANDARD; PRT; 180 AA.  
 AC P17333;

01-AUG-1990 (Rel. 15, Created)  
 01-AUG-1990 (Rel. 15, Last sequence update)  
 15-DEC-1998 (Rel. 37, Last annotation update)  
 Napin precursor (1.7S seed storage protein).  
 NAPI  
 Brassica napus (Rape).  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 NCBI\_TaxID=3708;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=cv. Topas;  
 RC STRAIN=cv. Topas; TISSUE=Leaf;  
 RX MEDLINE=91346654; PubMed=2102844;  
 RA Bazczynski C.L., Fallis L.;  
 RA "Isolation and nucleotide sequence of a genomic clone encoding a new  
 ET Brassica napus napin gene.";  
 RL Plant Mol. Biol. 14:633-635(1990).  
 [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN=cv. Topas;  
 RX MEDLINE=91346654; PubMed=2102844;  
 RA Bazczynski C.L., Iyer V.N., Miki B.L.;  
 RA Boutilier K.A., Gines M.J., Demoor J.M., Huang B.,  
 ET "Expression of the Rnmnap subfamily of napin genes coincides with the  
 RL induction of Brassica microspore embryogenesis.";  
 RL Plant Mol. Biol. 26:1711-1723(1994).  
 CC -!- FUNCTION: The small, basic, water-soluble napins are one of the  
 CC two major kinds of storage proteins synthesized in the seed during  
 CC its maturation.  
 CC -!- SUBUNIT: The mature protein consists of a small and a large chain  
 CC linked by disulfide bonds.  
 CC -!- TISSUE SPECIFICITY: Cotyledons and the axis.  
 CC -!- SIMILARITY: Belongs to the 2S seed storage albumins family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; X17542; CAA35580.1; -;  
 CC EMBL; U04945; AAA81909.1; -;  
 CC PIR; S10018; S10018.  
 CC InterPro; IPR003612; AAI.  
 CC InterPro; IPR000617; Napin.  
 CC Pfam; PF00234; tryd\_alpha\_amy1; 1.  
 CC PRINTS; PRO0496; NAPI.  
 CC ProDom; PD002498; Napin; 1.  
 CC SMART; SM00499; AAI; 1.  
 CC Seed storage protein; Signal; Multigene family.  
 KW SIGNAL 1 21  
 FT PROPEP 22 38  
 FT CHAIN 39 74 SMALL CHAIN.  
 FT PROPEP 75 94  
 FT CHAIN 95 180 LARGE CHAIN.  
 SQ SEQUENCE 180 AA; 20318 MW; 6F3883CBED55FE26 CRC64;  
 Query Match 70.0%; Score 35; DB 1; Length 180;  
 Best Local Similarity 55.8%; Pred. No. 7.3;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CSVPPLTSVC 9  
 Db 161 CNIPQSVSVC 169  
 RESULT 7  
 NKGD MACMU STANDARD; PRT; 216 AA.  
 ID NKGD MACMU  
 AC Q9WZJ7;  
 28-FEB-2003 (Rel. 41, Created)  
 28-FEB-2003 (Rel. 41, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 NKG2-D type II integral membrane protein (NKG2-D activating NK  
 DE receptor) (NK cell receptor D).  
 DE NKG2D.  
 GN NKG2D.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 NCBI\_TaxID=9544;  
 [1]  
 SEQUENCE FROM N.A.  
 RP MEDLINE=20322487; PubMed=10866118;  
 RA LaBonte M.L., Levy D.B., Letvin N.L.;  
 RA "Characterization of the novel transmembrane-deleted forms of NKG2-A, B, C,  
 RT identification of novel transmembrane-deleted forms of NKG2-A, B, C,  
 RT and D.";  
 RL Immunogenetics 51:496-499(2000).  
 CC -!- FUNCTION: Plays a role as a receptor for the recognition of MHC  
 CC class I HLA-E molecules by NK cells and some cytotoxic T-cells.  
 CC -!- SUBUNIT: Can form disulfide-bonded heterodimer with CD94.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein.  
 CC -!- TISSUE SPECIFICITY: Natural killer cells.  
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AF190943; AAF74539.1; -;  
 CC InterPro; IPR001304; Lectin\_C.  
 CC Pfam; PF00059; lectin\_c; 1.  
 CC SMART; SM00034; CLECT; 1.  
 CC PROSITE; PS00041; C-TYPE LECTIN 2; 1.  
 CC Receptor; Transmembrane; Multigene family; Signal-anchor; Lectin;  
 KW Glycoprotein; Polymorphism.  
 FT DOMAIN 1 51  
 FT TRANSMEM 52 72  
 FT SIGNAL-ANCHOR (POTENTIAL).  
 FT CYTOPLASMIC (POTENTIAL).  
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 73 216  
 FT DOMAIN 98 213  
 FT DISULFID 99 110  
 FT DISULFID 127 211  
 FT DISULFID 189 203  
 FT CARBOHYD 115 115  
 FT CARBOHYD 131 131  
 FT CARBOHYD 163 163  
 FT CARBOHYD 202 202  
 FT CARBOHYD 216 216  
 SQ SEQUENCE 216 AA; 25075 MW; A44883F31400DEAC CRC64;  
 Query Match 70.0%; Score 35; DB 1; Length 216;  
 Best Local Similarity 55.6%; Pred. No. 8.9;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CSVPPLTSVC 9  
 Db 203 CSIPNTYIC 211  
 RESULT 8  
 TYRO NEUCR STANDARD; PRT; 620 AA.  
 ID TYRO NEUCR  
 AC P00440;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Tyrosinase precursor (EC 1.14.18.1) (Monophenol monooxygenase).  
 GN T.

OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TS, and Oak Ridge.  
 RX MEDLINE=9000884; PubMed=2529259;  
 RA Kupper U., Niedermann D.M., Travaglini G., Lerch K.;  
 RT "Isolation and characterization of the tyrosinase gene from  
 RT Neurospora crassa."  
 RL J. Biol. Chem. 264:17250-17258 (1989).  
 RN (2)  
 RP SEQUENCE OF 1-407.  
 RC STRAIN=TL;  
 RX MEDLINE=82190018; PubMed=6210696;  
 RA Lerch K.;  
 RT "Primary structure of tyrosinase from Neurospora crassa. II. Complete  
 RT amino acid sequence and chemical structure of a tripeptide containing  
 RT an unusual thioether."  
 RL J. Biol. Chem. 257:6414-6419 (1982).  
 RN (3)  
 RP SEQUENCE OF 1-407.  
 RC STRAIN=TS, and Singi;  
 RX MEDLINE=82190019; PubMed=6210697;  
 RA Ruegg C., Ammer D., Lerch K.;  
 RT "Comparison of amino acid sequence and thermostability of tyrosinase  
 RT from three wild type strains of Neurospora crassa."  
 RL J. Biol. Chem. 257:6420-6426 (1982).  
 CC -!- FUNCTION: THIS IS A COPPER-CONTAINING OXIDASE THAT FUNCTIONS IN  
 CC THE FORMATION OF PIGMENTS SUCH AS MELANINS AND OTHER POLYPHENOLIC  
 CC COMPOUNDS.  
 CC -!- CATALYTIC ACTIVITY: L-tyrosine + L-DOPA + O(2) = L-DOPA +  
 CC DOPAquinone + H(2)O.  
 CC -!- COFACTOR: Binds 2 copper ions per subunit (By similarity).  
 CC -!- MISCELLANEOUS: THE STRAIN TS SEQUENCE IS SHOWN.  
 CC -!- SIMILARITY: Belongs to the tyrosinase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M32843; AAA33619.1; -;  
 CC EMBL; M3271; AAA33618.1; -;  
 CC InterPro; IPR008922; Di-copper\_centre.  
 CC Pfam; PF00264; tyrosinase; 1.  
 CC PRINTS; PR00092; TYROSINASE.  
 CC PROSITE; PS00497; TYROSINASE 1; 1.  
 CC PROSITE; PS00498; TYROSINASE 2; 1.  
 KW Melanin biosynthesis; Oxidoreductase; Monooxygenase; Copper;  
 KW acetylation; Thioether bond.  
 FT INIT MET 0  
 FT CHAIN 1 407  
 FT PROPEP 408 620  
 FT MOD RES 1 1  
 FT METAL 66 66  
 FT METAL 96 96  
 FT METAL 105 105  
 FT METAL 277 277  
 FT METAL 281 281  
 FT METAL 306 306  
 FT METAL 34 14  
 FT CROSSINK 94 96  
 FT VARIANT 14 14  
 FT VARIANT 29 29  
 FT VARIANT 130 130  
 FT VARIANT 345 346  
 FT VARIANT 370 370  
 FT VARIANT 423 423  
 FT VARIANT 423 423

FT VARIANT 449 449 K -> R (IN STRAIN OR).  
 FT CONFLICT 234 234 N -> D (IN REF. 2 AND 3).  
 SQ SEQUENCE 620 AA; 68546 MW; D12A30BB6A01D312 CRC64;  
 Query Match 70.0%; Score 35; DB 1; Length 620;  
 Best Local Similarity 75.0%; Pred. No. 27; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 2;  
 QY 2 SVPLTSVC 9  
 Db 589 TVPLTSLC 596  
 RESULT 9  
 ID SYR MOUSE STANDARD; PRT; 660 AA.  
 AC Q9D019; Q8VDW1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).  
 GN PARS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,  
 RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690 (2001).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Sheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

```

CC CC -1- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC CC diphosphate + L-arginyl-tRNA(Arg).
CC CC -1- SUBUNIT: Monomer; also part of a multisubunit complex that groups
CC CC tRNA ligases for Arg, Asp, Glu, Gln, Ile, Leu, Lys, Met and Pro.
CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -1- DOMAIN: The N-terminal (AA 1-72) has two regions predicted to be
CC CC alpha-helical that might be involved in the multisynthetase
CC CC complex assembly.
CC CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AK011383; BAB27583.1; -.
CC CC EMBL; BC020132; AAH20132.1; -.
CC CC MGD; MGI:1914297; Rars.
CC CC InterPro; IPR001278; Arg tRNA-synt_1c.
CC CC InterPro; IPR005148; N.
CC CC InterPro; IPR008909; tRNA-synt_1d_C.
CC CC InterPro; IPR001412; tRNA-synt_1.
CC CC Pfam; PF03485; N-Arg; 1.
CC CC Pfam; PF00750; tRNA-synt_1d; 1.
CC CC Pfam; PF05746; tRNA-synt_1d_C; 1.
CC CC PRINTS; PR01038; TRNASYNTHARG.
CC CC TIGRFAMs; TIGR00456; argS; 1.
CC CC PROSITE; PS00178; AA tRNA-LIGASE_I; 1.
CC CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
CC CC KW DOMAIN 1 72
CC CC COULD BE INVOLVED IN THE ASSEMBLY OF THE
CC CC MULTISYNTHETASE COMPLEX.
CC CC SITE 201 212
CC CC "HIGH" REGION.
CC CC L -> H (IN REF. 1).
CC CC G -> V (IN REF. 1).
CC CC M -> L (IN REF. 1).
CC CC CONFLICT 474 474
CC CC CONFLICT 474 474
CC CC SEQUENCE 660 AA; 75673 MW; 6E8EBCC590FAB81D CRC64;

Query Match 70.0%; Score 35; DB 1; Length 660;
Best Local Similarity 87.5%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CSVPLTVC 8
Db 369 CSVPLTIV 376

RESULT 10
ALL1 BRAJU STANDARD; PRT; 129 AA.
AC P80207; P80215;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 42, Last annotation update)
DE Allergen Bra j 1-E, small and large chains (Bra j 1).
OS Brassica juncea (Leaf mustard) (Indian mustard).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Brassica.
OC NCBI_TaxID=3707;
OX [1]
RN RP SEQUENCE.
RC MEDLINE=93356721; PubMed=7688955;
RX Monsalve R.I., Gonzalez de la Pena M.A., Menendez-Atlas L.,
RA Lopez-Otin C., Villalba M., Rodriguez R.;
RT "Characterization of a new oriental-mustard (Brassica juncea)
allergen, Bra j 1E: detection of an allergenic epitope.";
RL Biochem. J. 293:625-632(1993).
CC -1- FUNCTION: This is a 2S seed storage protein.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN

```

```

CC CC LINKED BY TWO DISULFIDE BONDS.
CC CC -1- ALLERGEN: Causes an allergic reaction in human.
CC CC -1- SIMILARITY: Belongs to the 2S seed storage albumins family.
CC CC PIR; S35591; S35591.
CC CC InterPro; IPR003612; AAI.
CC CC InterPro; IPR000617; Napin.
CC CC PRINTS; PR00496; NAPIN.
CC CC ProDom; PD002498; Napin; 1.
CC CC SMART; SM00499; AAI; 1.
CC CC Allergen; Seed storage protein.
CC CC KW CHAIN 1 37 SMALL CHAIN.
CC CC FT NON CONS 37 38
CC CC FT CHAIN 38 129 LARGE CHAIN.
CC CC FT VARIANT 6 6 F -> I.
CC CC FT VARIANT 20 20 R -> K.
CC CC SEQUENCE 129 AA; 14644 MW; D6F28E03F62B08F8 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 129;
Best Local Similarity 44.4%; Pred. No. 8.1;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPLTVC 9
Db 111 CNIPRVVIC 119

RESULT 11
2SSB BRANA STANDARD; PRT; 178 AA.
AC P27740;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Napin B precursor (1.7S seed storage protein).
DE NAPP.
GN NAPP.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Brassica.
OC NCBI_TaxID=3708;
OX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Svalofsk Karat 20516-K;
RX MEDLINE=91231016; PubMed=2029903;
RA Ericson M.L., Muren E., Gustavsson H.O., Josefsson L.G., Rask L.;
RT "Analysis of the promoter region of napin genes from Brassica napus
demonstrates binding of nuclear protein in vitro to a conserved
sequence motif.";
RL Eur. J. Biochem. 197:741-746(1991).
CC -1- FUNCTION: The small, basic, water-soluble napins are one of the
CC CC two major kinds of storage proteins synthesized in the seed during
CC CC its maturation.
CC CC -1- SUBUNIT: The mature protein consists of a small and a large chain
CC CC linked by disulfide bonds.
CC CC -1- TISSUE SPECIFICITY: Cotyledons and the axis.
CC CC -1- SIMILARITY: Belongs to the 2S seed storage albumins family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X58142; CAA41150.1; -.
CC CC PIR; S15382; S15382.
CC CC InterPro; IPR003612; AAI.
CC CC InterPro; IPR000617; Napin.
CC CC Pfam; PF00234; tryp_alpha_amyl; 1.
CC CC PRINTS; PR00496; NAPIN.
CC CC ProDom; PD002498; Napin; 1.
CC CC SMART; SM00499; AAI; 1.

```

KW Seed storage protein; Signal; Multigene family; Embryo.  
 FT SIGNAL 1 21 BY SIMILARITY.  
 FT PROPEP 22 38 BY SIMILARITY.  
 FT CHAIN 39 74 SMALL CHAIN (BY SIMILARITY).  
 FT PROPEP 75 94 BY SIMILARITY.  
 FT CHAIN 95 178 LARGE CHAIN (BY SIMILARITY).  
 SQ SEQUENCE 178 AA; 20114 MW; 96CE0ADB7CD966E9 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 178;  
 Best Local Similarity 55.6%; Pred. No. 11;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CSVPVLTVC 9  
 Db 159 CKIPQVVC 167

RESULT 12  
 ID\_EPA2\_RALSO STANDARD; PRT; 377 AA.  
 AC Q45407;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE EPS I polysaccharide export outer membrane protein epsA precursor.  
 GN EPSA.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AW;  
 RX MEDLINE=96059643; PubMed=7476194;  
 RA Huang J., Schell M.;  
 RT "Molecular characterization of the eps gene cluster of Pseudomonas  
 solanacearum and its transcriptional regulation at a single  
 promoter.";  
 RT Mol. Microbiol. 16:977-989(1995).  
 RL CC -1- FUNCTION: PROBABLY INVOLVED IN POLYMERIZATION AND/OR EXPORT OF  
 CC EXOPOLYSACCHARIDE EPS I WHICH FUNCTIONS AS A VIRULENCE FACTOR.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE BEXD/CTRA/VEXA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U17898; AAA91624.1; --  
 DR PIR; S77634; S77634.  
 DR InterPro; IPR003715; Poly.export.  
 DR InterPro; IPR000437; Prok lipoprot\_s.  
 DR Pfam; PF02563; Poly.export; 1.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; FALSE\_NEG.  
 KW Polysaccharide transport; Transport; Outer membrane; Transmembrane;  
 KX Lipoprotein; Porin; Signal; Palmitate.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 377 EPS I POLYSACCHARIDE EXPORT OUTER  
 FT LIPID 24 24 MEMBRANE PROTEIN EPSA.  
 FT FT S-diacylglycerol cysteine (By  
 FT FT similarity).  
 FT LIPID 24 24 N-palmitoyl cysteine (By similarity).  
 SQ SEQUENCE 377 AA; 40446 MW; BE0084252E819309 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 377;  
 Best Local Similarity 55.6%; Pred. No. 25;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPVLTVC 9  
 Db 16 CAVPLMAAC 24

RESULT 13  
 ID\_P12R\_HUMAN STANDARD; PRT; 386 AA.  
 AC P43119;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Prostaglandin receptor (Prostanoid IP receptor) (PGI receptor)  
 DE (Prostaglandin I2 receptor).  
 GN PTGIR OR PRIPR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=94216334; PubMed=7512962;  
 RA Boie Y., Rushmore T.H., Darmon-Goodwin A., Grygorczyk R.,  
 RA Slipetz D.M., Metters K.M., Abramovitz M.;  
 RT "Cloning and expression of a cDNA for the human prostanoid IP  
 receptor.";  
 RT J. Biol. Chem. 269:12173-12178(1994).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94237286; PubMed=7514139;  
 RA Katsuyama M., Sugimoto Y., Namba T., Irie A., Negishi M.,  
 RA Narumiya S., Ichikawa A.;  
 RT "Cloning and expression of a cDNA for the human prostacyclin  
 receptor.";  
 RT FEBS Lett. 344:74-78(1994).  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=95008086; PubMed=7923647;  
 RA Nakagawa O., Tanaka I., Usui T., Harada M., Sasaki Y., Itoh H.,  
 RA Yoshimasa T., Namba T., Narumiya S., Nakao K.;  
 RT "Molecular cloning of human prostacyclin receptor cDNA and its gene  
 expression in the cardiovascular system.";  
 RT Circulation 90:1643-1647(1994).  
 RL [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95394450; PubMed=7665161;  
 RA Ogawa Y., Tanaka I., Inoue M., Yoshitake Y., Isse N., Nakagawa O.,  
 RA Usui T., Itoh H., Yoshimasa T., Narumiya S.;  
 RT "Structural organization and chromosomal assignment of the human  
 prostacyclin receptor gene.";  
 RT Genomics 27:142-148(1995).  
 RL [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Warren C.N., Aronstam R.S., Sharma S.V.;  
 RT "cDNA clones of human proteins involved in signal transduction  
 RT sequenced by the Guthrie cDNA resource center ([www.cdna.org](http://www.cdna.org)).";  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP PALMITOYLATION.  
 RX MEDLINE=22486644; PubMed=12488443;  
 RA Miggin S.M., Lawler O.A., Kinsella B.T.;  
 RT "Palmitoylation of the human prostacyclin receptor. Functional  
 RT implications of palmitoylation and isoprenylation.";  
 RL J. Biol. Chem. 278:6947-6958(2003).  
 CC -1- FUNCTION: Receptor for prostacyclin (prostaglandin I2 or PGI2).  
 CC The activity of this receptor is mediated by G(s) proteins which  
 CC activate adenylate cyclase.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; L29016; AAA36448.1; --  
EMBL; D25418; BAA05008.1; --  
EMBL; D29634; BAA06110.1; --  
EMBL; D38127; BAA07325.1; --  
EMBL; D38128; BAA07325.1; JOINED.  
EMBL; AY242134; AA092301.1; --  
PIR; A57066; A57066.  
Genew; HGNC:9602; PTGIR.  
MIM; 600224; --  
GO; GO:0005887; C: integral to plasma membrane; TAS.  
GO; GO:0004959; F: prostaglandin I receptor activity; TAS.  
GO; GO:0007267; P: cell-cell signaling; TAS.  
GO; GO:0007187; P: G-protein signaling, coupled to cyclic nucl. . .; TAS.  
InterPro; IPR000276; GPCR Rhodopsin.  
InterPro; IPR008365; ProstanoidR.  
Pfam; PF00001; 7tm.1; 1.  
PRINTS; PR00237; GPCRHDOPSN.  
PRINTS; PR01788; PROSTANOIDR.  
PROSITE; PS00237; G-PROTEIN RECP F1.1; 1.  
PROSITE; PS0262; G-PROTEIN RECP F1.2; 1.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Lipoprotein;  
Palmitate.  
FT DOMAIN 1 16 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 17 38 1 (POTENTIAL).  
FT DOMAIN 39 51 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 52 76 2 (POTENTIAL).  
FT DOMAIN 77 94 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 95 115 3 (POTENTIAL).  
FT DOMAIN 116 134 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 135 158 4 (POTENTIAL).  
FT DOMAIN 159 181 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 182 208 5 (POTENTIAL).  
FT DOMAIN 209 235 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 236 260 6 (POTENTIAL).  
FT DOMAIN 261 274 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 275 295 7 (POTENTIAL).  
FT DOMAIN 296 386 CYTOPLASMIC (POTENTIAL).  
FT LIPID 308 308 S-palmitoyl cysteine.  
FT LIPID 311 311 N-linked (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 7 7 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 386 AA; 40955 MW; 2B680CDBACED1608 CRC64;  
Query Match 68.0%; Score 34; DB 1; Length 386;  
Best Local Similarity 66.7%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CSVPLTSVC 9  
Db 251 CSLPLTIRC 259  
RESULT 14  
ID\_CIKS\_HUMAN STANDARD; PRT; 574 AA.  
AC Q43734; Q9H5W2; Q9H6Y3; Q9NS14; Q9UG72;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Adapter protein CIKS (Connection to IKK and SAPK/JNK) (Nuclear factor  
DE NF-kappa-B activator 1) (ACT1).  
GN C6ORF4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;

[1] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RP MEDLINE=20363742; PubMed=10903453;  
RX Morelli C., Magnanini C., Mungall A.J., Negrini M.,  
RA Barabanti-Brodano G.;  
RT "Cloning and characterization of two overlapping genes in a subregion  
at 6q21 involved in replicative senescence and schizophrenia.";  
RL Gene 252:217-225(2000).  
[2]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RP TISSUE=Embryonic Kidney;  
RX MEDLINE=20442402; PubMed=10962024;  
RA Li X., Commane M., Nie H., Hua X., Chatterjee-Kishore M., Wald D.,  
RA Haag M., Stark G.R.;  
RT "Act1, an NF-kappa B-activating protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:10489-10493(2000).  
[3]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RP TISSUE=Liver;  
RX MEDLINE=20442403; PubMed=10962033;  
RA Leonardi A., Chariot A., Claudio E., Cunningham K., Siebenlist U.;  
RT "CIKS, a connection to Ikappa B kinase and stress-activated protein  
kinase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:10494-10499(2000).  
[4]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RP Palmer S.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
[5]  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RP TISSUE=Uterus;  
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
[6]  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RP TISSUE=Colon;  
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,  
RA Nakamura Y., Isoigai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
[7]  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RP TISSUE=Lymph;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: Could be involved in the activation of both NF-kappa-B  
via a NF-kappa-B inhibitor kinase (IKK)-dependent mechanism and  
stress-activated protein kinase (SAPK)/JNK.  
CC -!- SUBUNIT: Interacts with IKKGI/NF-kappa B essential modulator, with  
CHUK/IKK-alpha and with IKKBI/IKK-beta.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1; Synonyms=C6ORF4;



```

CC      IsoId=O43734-1; Sequence=Displayed;
CC      Name=2; Synonyms=C6ORF5, C6ORF6;
CC      IsoId=O43734-2; Sequence=VSP_004163;
CC      TISSUE SPECIFICITY: Widely expressed.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF136405; AAF67445.1; -
CC      EMBL; AF136406; AAF67446.1; -
CC      EMBL; AF136407; AAF67447.1; -
CC      EMBL; AF274303; AAG15367.1; -
CC      EMBL; AF272151; AAG15407.1; -
CC      EMBL; AL008730; CAA15506.1; -
CC      EMBL; AL008730; CAA15507.1; ALT_INIT.
CC      EMBL; AL050289; CAB43390.1; -
CC      EMBL; AK025351; BAB15117.1; -
CC      EMBL; AK026602; BAB15507.1; ALT_INIT.
CC      EMBL; BC002823; AAH02823.1; -
CC      PIR; T08794; T08794.
CC      Genew; HGNC:1343; C6orf4.
CC      MIM; 607043; -.
CC      GO; GO:0007242; P:intracellular signaling cascade; NAS.
CC      KW Alternative splicing.
CC      DOMAIN 135 189 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC      VARSPLIC 1 9 Missing (in isoform 2).
CC      FT CONFLICT 332 332 Q -> H (IN REF. 1; AAF67445, 2 AND 4).
CC      FT CONFLICT 334 334 E -> D (IN REF. 3).
CC      FT CONFLICT 347 347 P -> S (IN REF. 1; AAF67447).
CC      FT CONFLICT 463 463 MISSING (IN REF. 6; BAB15117).
CC      SQ SEQUENCE 574 AA; 64657 MW; E3957DEA5E349094 CRC64;
CC
CC      Query Match 68.0%; Score 34; DB 1; Length 574;
CC      Best Local Similarity 85.7%; Pred. No. 40;
CC      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC      QY 3 VPLTSVC 9
CC      Db 225 LPLTSVC 231
CC
CC      RESULT 15
CC      ID_CSP1_RAT STANDARD; PRT; 586 AA.
CC      AC Q9ULH6;
CC      DT 10-OCT-2003 (Rel. 42, Created)
CC      DT 10-OCT-2003 (Rel. 42, Last sequence update)
CC      DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC      DE CDK5 regulatory subunit associated protein 1 (CDK5 activator-binding
CC      DE protein C42).
CC      GN CDK5RAP1.
CC      OS Rattus norvegicus (Rat).
CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC      OX NCBI_TaxID=10116;
CC      RN [1]_TaxID=10116;
CC      RP SEQUENCE FROM N.A., AND INTERACTION WITH CDK5R1.
CC      RC TISSUE=Brain;
CC      RX MEDLINE=20184747; PubMed=10721722;
CC      RA Ching Y.-P., Qi Z., Wang J.H.;
CC      RT "Cloning of three novel neuronal Cdk5 activator binding proteins.";
CC      RL Gene 242:285-294(2000).
CC      RN [2]
CC      RP INTERACTION WITH CDK5R1.
CC      RX MEDLINE=21975289; PubMed=11882646;
CC      RA Ching Y.-P., Pang A.S.H., Lam W.-H., Qi R.Z., Wang J.H.;
CC      RT "Identification of a neuronal Cdk5 activator-binding protein as Cdk5

```

```

RT inhibitor";
RL J. Biol. Chem. 277:15237-15240(2002).
CC      FUNCTION: Probable regulator of CDK5 activity. May inhibit CDK5
CC      function via its interaction with CDK5R1.
CC      SUBUNIT: Interacts with CDK5R1 and CDK5R2. The interaction with
CC      CDK5R1 is prevented by the association between CDK5R1 and
CC      CDK5RAP2. Interacts also with the complex CDK5R1-CDK5. Does not
CC      interact with monomeric CDK5. The interaction is prevented by the
CC      association between CDK5R1 and CDK5RAP2.
CC      DOMAIN: The C-terminal part (475-586) is necessary for the
CC      interaction with CDK5R1, while the N-terminal part (1-168) is
CC      required for inhibiting the activity of the CDK5 kinase.
CC      SIMILARITY: Belongs to the UPF0004 family. CDK5RAP1 subfamily.
CC      SIMILARITY: Contains 1 TRAM domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF177477; AAF60223.1; -
CC      GO; GO:0042808; P:neuronal Cdc2-like kinase binding; ISS.
CC      GO; GO:0007420; P:brain development; ISS.
CC      GO; GO:0008283; P:cell proliferation; NAS.
CC      GO; GO:0045736; P:negative regulation of CDK activity; ISS.
CC      GO; GO:0045664; P:regulation of neuron differentiation; ISS.
CC      InterPro; IPR006638; ELP3.
CC      InterPro; IPR006463; MiA_methiolase.
CC      InterPro; IPR007197; Radical_SAM.
CC      InterPro; IPR002792; TRAM.
CC      InterPro; IPR005839; UPF0004.
CC      Pfam; PF04055; Radical_SAM; 1.
CC      Pfam; PF01938; TRAM; 1.
CC      Pfam; PF00919; UPF0004; 1.
CC      SMART; SM00729; ELP3; 1.
CC      TIGRFAMS; TIGR01574; miA-methiolase; 1.
CC      TIGRFAMS; TIGR00089; TIGR00089; 1.
CC      PROSITE; PS0926; TRAM; 1.
CC      PROSITE; PS01278; UPF0004; 1.
CC      DOMAIN 498 573
CC      SQ SEQUENCE 586 AA; 65866 MW; 95AC432AA3FF552C CRC64;
CC
CC      Query Match 68.0%; Score 34; DB 1; Length 586;
CC      Best Local Similarity 75.0%; Pred. No. 41;
CC      Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC      QY 2 SVPLTSVC 9
CC      Db 14 SAPLTSVC 21
CC
CC      Search completed: September 5, 2004, 09:56:07
CC      Job time : 6.09091 secs

```

**This Page Blank (uspto)**

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:46:09 ; Search time 20.3636 Seconds  
(without alignments)  
139.448 Million cell updates/sec

Title: US-09-761-636A-10  
Perfect score: 50  
Sequence: 1 CSVPLTVC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	86.0	1487	11	Q62028 mus musculu
2	41	82.0	349	16	Q82LE3 salmonella
3	41	82.0	349	16	Q82258 salmonella
4	41	82.0	349	16	Q83J80 shigella fl
5	41	82.0	376	16	Q8FCN4 escherichia
6	41	82.0	409	11	Q63791 rattus sp.
7	41	82.0	585	10	Q8RXR2 arabidopsis
8	41	82.0	750	10	Q9ARC9 lycopersico
9	41	82.0	750	10	Q9ARC6 lycopersico
10	41	82.0	780	10	Q9ARF2 arabidopsis
11	41	82.0	781	10	Q9ARM3 arabidopsis
12	41	82.0	784	10	Q9LG04 arabidopsis
13	39	78.0	432	17	Q58874 methanococc
14	38	76.0	323	4	Q8WVW3 homo sapien
15	38	76.0	491	4	Q9H6R1 homo sapien
16	38	76.0	786	4	Q9H747 homo sapien

17	38	76.0	802	4	Q8TB22
18	37	74.0	97	16	Q8CLM0
19	37	74.0	247	6	Q8HZR8
20	37	74.0	718	10	Q94JF5
21	37	74.0	1977	5	Q9V6A0
22	36	72.0	124	11	Q9D179
23	36	72.0	168	4	Q96PA7
24	36	72.0	201	4	Q9H1K3
25	36	72.0	247	4	Q9BXN2
26	36	72.0	591	10	Q39485
27	36	72.0	1217	4	Q9ULF5
28	36	72.0	2192	5	Q01768
29	35	70.0	86	10	Q9S9F0
30	35	70.0	88	10	Q9S9E7
31	35	70.0	88	10	Q9S9E9
32	35	70.0	88	10	Q9S9E8
33	35	70.0	91	16	Q97HY9
34	35	70.0	125	10	P80208
35	35	70.0	145	10	Q41278
36	35	70.0	145	10	Q41281
37	35	70.0	145	10	Q41280
38	35	70.0	145	10	Q41279
39	35	70.0	155	10	Q41167
40	35	70.0	168	5	Q9NL63
41	35	70.0	178	10	Q93944
42	35	70.0	178	10	Q96339
43	35	70.0	178	10	Q42491
44	35	70.0	178	10	Q42490
45	35	70.0	178	10	Q42413

## ALIGNMENTS

## RESULT 1

Q62028 PRELIMINARY; PRT; 1487 AA.  
AC Q62028;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Phospholipase A2 receptor precursor.  
GN PLA2GLBR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95010128; PubMed=7925459;  
RA Higashino K., Ishizaki J., Kishino J., Ohara O., Arita H.;  
RT "Structural comparison of phospholipase-A2-binding regions in  
RT phospholipase-A2 receptors from various mammals.";  
RL Eur. J. Biochem. 225:375-382(1994).  
DR EMBL; D30779; BAA06443.1; -;  
DR PIR; S48719; S48719.  
DR HSSP; P02751; 2FN2.  
DR MGD; MGI:102468; Pla2glbr.  
DR GO; GO:0004872; F:receptor activity; IEA.  
DR GO; GO:0005529; F:sugar binding; IEA.  
DR InterPro; IPR000562; FN\_Type\_II.  
DR InterPro; IPR001304; LECTIN\_C.  
DR InterPro; IPR008997; RicinB\_like.  
DR InterPro; IPR000772; Ricin\_B\_lectin.  
DR Pfam; PF00040; fn2; 1.  
DR Pfam; PF00059; lectin.c; 8.  
DR PRINTS; PR00013; FNTYPEII.  
DR ProDom; PD000395; FN\_Type\_II; 1.  
DR SMART; SM00034; CLECT; 8.  
DR SMART; SM00059; FN2; 1.  
DR SMART; SM00458; RICIN; 1.  
DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 2.  
DR PROSITE; PS00041; C\_TYPE\_LECTIN\_2; 8.

Q8cb22 homo sapien  
Q8clm0 versinia pe  
Q8hzr8 macaca mula  
Q94jf5 oryza sativ  
Q9v6a0 drosophila  
Q9d179 mus musculu  
Q96pa7 homo sapien  
Q9h1k3 homo sapien  
Q9bxn2 homo sapien  
Q39485 chlamydomon  
Q9ull5 homo sapien  
Q01768 caenorhabdi  
Q9s9f0 brassica na  
Q9s9e7 brassica na  
Q9s9e9 brassica na  
Q9s9e8 brassica na  
Q97hy9 clostridium  
P80208 brassica na  
Q41278 sinapis alb  
Q41281 sinapis alb  
Q41280 sinapis alb  
Q41279 sinapis alb  
Q41167 raphanus sa  
Q9nl63 sarcophaga  
Q93944 brassica na  
Q96339 brassica ca  
Q42491 brassica ni  
Q42490 brassica ol  
Q42413 brassica ju

```

DR PROSITE; PS00023; FIBRONECTIN 2; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Receptor; Signal.
FT CHAIN 27 1487 PHOSPHOLIPASE A2 RECEPTOR.
SQ SEQUENCE 1487 AA; 170511 MW; AD8D905859B0EDE8 CRC64;

Query Match
Best Local Similarity 86.0%; Score 43; DB 11; Length 1487;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9
   |||||:|
Db 929 CSVPLPSIC 937

RESULT 2
ID Q8ZLE3 PRELIMINARY; PRT; 349 AA.
AC Q8ZLE3;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative PerM family permease.
GN YHHT OR STM3582.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856 (2001).
DR EMBL; AE008865; AL22442.1; -.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 349 AA; 38301 MW; 21A55D9F2C5FDB9A CRC64;

Query Match
Best Local Similarity 82.0%; Score 41; DB 16; Length 349;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVPLTSVC 9
   |||||
Db 314 SVPLTSVC 321

RESULT 3
ID Q8Z258 PRELIMINARY; PRT; 349 AA.
AC Q8Z258;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Putative membrane protein.
GN STY4229 OR T3940.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;

Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
Buriland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18."
RL J. Bacteriol. 185:2330-2337 (2003).
DR EMBL; AL627281; CAD08048.1; -.
DR EMBL; AE016847; AA071411.1; -.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 349 AA; 38359 MW; 226A5C802E8FDB9A CRC64;

Query Match
Best Local Similarity 82.0%; Score 41; DB 16; Length 349;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SVPLTSVC 9
   |||||
Db 314 SVPLTSVC 321

RESULT 4
ID Q83J80 PRELIMINARY; PRT; 349 AA.
AC Q83J80;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Orf, conserved hypothetical protein (Hypothetical protein yhhT).
GN YHHT OR SF3492 OR S4271.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
Sun L., Xue Y., Zhao A., Gao Y., Kan B., Ding K., Chen S.,
Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL; AE015357; AAN44951.1; -.

```

```

DR EMBL; AE016992; AAP19231.1; -.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 349 AA; 38464 MW; 935ADBA85BD5651 CRC64;

Query Match      82.0%; Score 41; DB 16; Length 349;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SVPLTSVC 9
Db      314 SVPLTSVC 321

RESULT 5
Q8FCN4
ID Q8FCN4 PRELIMINARY; PRT; 376 AA.
AC Q8FCN4
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein yhhT.
GN yhhT OR C4267.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=O6:HL / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackert J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016768; AAN82703.1; -.
DR InterPro; IPR002549; UPF0118.
DR Pfam; PF01594; UPF0118; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 376 AA; 41438 MW; 6B39C765A5D7315A CRC64;

Query Match      82.0%; Score 41; DB 16; Length 376;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 SVPLTSVC 9
Db      341 SVPLTSVC 348

RESULT 6
Q63791
ID Q63791 PRELIMINARY; PRT; 409 AA.
AC Q63791
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospholipase A2 receptor (fragment).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Smooth muscle;
RX MEDLINE=95010128; PubMed=7925459;
RA Higashino K., Ishizaki J., Kishino J., Ohara O., Arita H.;
RT "Structural comparison of phospholipase-A2-binding regions in
RT phospholipase-A2 receptors from various mammals.";

```

```

RL Eur. J. Biochem. 225:375-382(1994).
DR EMBL; D30781; BAA06445.1; -.
DR HSP; P23807; IIXX.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_C; 3.
DR SMART; SM00034; CLECT; 3.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 3.
KW Receptor.
FT NON_TER 1
FT NON_TER 409
SQ SEQUENCE 409 AA; 47740 MW; 56D957D2DCA00AD8 CRC64;

Query Match      82.0%; Score 41; DB 11; Length 409;
Best Local Similarity 66.7%; Pred. No. 5.1;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CSVPLTSVC 9
Db      396 CSVPLTSVC 404

RESULT 7
Q8RXR2
ID Q8RXR2 PRELIMINARY; PRT; 585 AA.
AC Q8RXR2
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN Atlg55620.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis Full length cDNA Clones.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY080722; AAL86324.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.
DR GO; GO:0006821; P:chloride transport; IEA.
DR InterPro; IPR000644; CBS domain.
DR InterPro; IPR001807; Cl-Channel_volt.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF00654; voltage_CLC; 1.
DR PRINTS; PR00762; CLCHANNEL.
DR SMART; SM00116; CBS; 2.
KW Hypothetical protein.
SQ SEQUENCE 585 AA; 62547 MW; E29CE584B81D826 CRC64;

Query Match      82.0%; Score 41; DB 10; Length 585;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CSVPLTSVC 8
Db      318 CSVPLTSVC 325

RESULT 8
Q9ARC9
ID Q9ARC9 PRELIMINARY; PRT; 750 AA.

```

KW Hypothetical protein.  
SQ SEQUENCE 750 AA; 79985 MW; 9286ED1EE2B5FCD4 CRC64;  
Query Match 82.0%; Score 41; DB 10; Length 750;  
Best Local Similarity 100.0%; Pred. No. 8.8; 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CSVPLTSV 8  
Db 474 CSVPLTSV 481  
RESULT 10  
Q9ARF2 PRELIMINARY; PRT; 780 AA.  
AC Q9ARF2; 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Capsella rubella.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Capsella.  
OX NCBI\_TaxID=81985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21178822; PubMed=11283350;  
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
RA Schumacher K., Schmitt G., Schmidt R.;  
RT "Comparative sequence analysis reveals extensive microcolinearity in  
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
RT genomes.";  
RL Plant Cell 13:979-988(2001).  
DR EMBL; AJ303349; CAC36391.1;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.  
DR GO; GO:0006821; P:chloride transport; IEA.  
DR InterPro; IPR000644; CBS domain.  
DR InterPro; IPR001807; CL-channel\_volt.  
DR Pfam; PF00571; CBS; 2.  
DR Pfam; PF00654; voltage CLC; 1.  
DR PRINTS; PR00762; CLCHANNEL.  
DR SMART; SM00116; CBS; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 780 AA; 83811 MW; 5718FA7E2AEE81FB CRC64;  
Query Match 82.0%; Score 41; DB 10; Length 780;  
Best Local Similarity 100.0%; Pred. No. 9.2; 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CSVPLTSV 8  
Db 513 CSVPLTSV 520  
RESULT 11  
Q9ARM3 PRELIMINARY; PRT; 781 AA.  
AC Q9ARM3; 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein (CLC-f chloride channel protein).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21178822; PubMed=11283350;  
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,

Q9ARC9;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21178822; PubMed=11283350;  
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
RA Schumacher K., Schmitt G., Schmidt R.;  
RT "Comparative sequence analysis reveals extensive microcolinearity in  
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
RT genomes.";  
RL Plant Cell 13:979-988(2001).  
DR EMBL; AJ303344; CAC36398.1;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.  
DR GO; GO:0006821; P:chloride transport; IEA.  
DR InterPro; IPR000644; CBS domain.  
DR InterPro; IPR001807; CL-channel\_volt.  
DR Pfam; PF00571; CBS; 2.  
DR Pfam; PF00654; voltage CLC; 1.  
DR PRINTS; PR00762; CLCHANNEL.  
DR SMART; SM00116; CBS; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 750 AA; 79994 MW; 400448B3D37042C CRC64;  
Query Match 82.0%; Score 41; DB 10; Length 750;  
Best Local Similarity 100.0%; Pred. No. 8.8; 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CSVPLTSV 8  
Db 474 CSVPLTSV 481  
RESULT 9  
Q9ARC6 PRELIMINARY; PRT; 750 AA.  
AC Q9ARC6;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21178822; PubMed=11283350;  
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,  
RA Schumacher K., Schmitt G., Schmidt R.;  
RT "Comparative sequence analysis reveals extensive microcolinearity in  
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
RT genomes.";  
RL Plant Cell 13:979-988(2001).  
DR EMBL; AJ303345; CAC36403.1;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.  
DR GO; GO:0006821; P:chloride transport; IEA.  
DR InterPro; IPR000644; CBS domain.  
DR InterPro; IPR001807; CL-channel\_volt.  
DR Pfam; PF00571; CBS; 2.  
DR Pfam; PF00654; voltage CLC; 1.  
DR PRINTS; PR00762; CLCHANNEL.  
DR SMART; SM00116; CBS; 2.

RA Schumacher K., Schmitz G., Schmidt R.;  
RT "Comparative sequence analysis reveals extensive microcolinearity in  
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella  
RT genomes.";  
RL Plant Cell 13:979-988(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Seeding aerial parts;  
RA Vinauger-Douard M., Charon C., Lapous D., Allot M., Granier F.,  
RA Bouchez D., Barbier-Brygoo H., Ephritikhine G.;  
RT "Molecular and functional characterization of AtCLC-f, a putative new  
RT Arabidopsis chloride channel.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ303348; CAC36386.1; -;  
DR EMBL; AF366368; AAKS3391.1; -;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.  
DR GO; GO:0006821; P:chloride transport; IEA.  
DR InterPro; IPR000644; C:domain.  
DR Pfam; PF00571; CBS; 2.  
DR PRINTS; PR00762; CLCHANNEL.  
DR SMART; SM00116; CBS; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 781 AA; 83548 MW; BE9DEB3603D9E0D8 CRC64;  
Query Match 82.0%; Score 41; DB 10; Length 781;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CSVPLTSV 8  
Db 514 CSVPLTSV 521  
RESULT 12  
Q9LGG04  
ID Q9LGG04 PRELIMINARY; PRT; 784 AA.  
AC Q9LGG04;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE F20N2.5.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsy N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome  
RT 1.";  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Shinn P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,  
RA Walker M.M., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,

RA Gonzalez A.A., Hansen N.N.F., Huizar L.L., Kremenetskaia I.I.,  
RA Lenz C.C., Li J.J., Liu S.S., Luros S.S., Rowley D.D., Schwartz J.J.,  
RA Toriumi M.M., Vysotskaia V.V., Yu G.G., Davis R.R.W.,  
RA Federspiel N.N.A., Theologis A.A., Ecker J.J.R.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC002328; AAF79509.1; -;  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005247; F:voltage-gated chloride channel activity; IEA.  
DR GO; GO:0006821; P:chloride transport; IEA.  
DR InterPro; IPR000644; C:domain.  
DR InterPro; IPR001807; CL-channel\_volt.  
DR Pfam; PF00571; CBS; 2.  
DR Pfam; PF00654; voltage CLC; 1.  
DR PRINTS; PR00762; CLCHANNEL.  
DR SMART; SM00116; CBS; 2.  
SQ SEQUENCE 784 AA; 83614 MW; 0495177BDA19330D CRC64;  
Query Match 82.0%; Score 41; DB 10; Length 784;  
Best Local Similarity 100.0%; Pred. No. 9.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CSVPLTSV 8  
Db 514 CSVPLTSV 521  
RESULT 13  
Q58874  
ID Q58874 PRELIMINARY; PRT; 432 AA.  
AC Q58874;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein MJ1479.  
GN MJ1479.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2861 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -!- SIMILARITY: TO E.COLI HYPOTHETICAL 45.5 KDA PROTEIN IN LRHA  
CC 3'REGION, H.INFLUENZAE HI0286 AND M.TUBERCULOSIS MTC279.04C.  
DR EMBL; U67589; AAB99491.1; -;  
DR PIR; F64484; F64484.  
DR TIGR; MJ1479; -;

DR GO: GO:0005622; C:intracellular; IEA.  
DR GO: GO:0005840; C:ribosome; IEA.  
DR GO: GO:0016947; F:1-aminocyclopropane-1-carboxylate synthase . . . ; IEA.  
DR GO: GO:0003735; F:structural constituent of ribosome; IEA.  
DR GO: GO:0008483; F:transaminase activity; IEA.  
DR GO: GO:0009058; P:biosynthesis; IEA.  
DR GO: GO:0006412; P:protein biosynthesis; IEA.  
DR InterPro: IPR001176; ACC\_synthase.  
DR InterPro: IPR004839; Aminotrans\_I/II.  
DR InterPro: IPR000529; Ribosomal\_S6.  
DR Pfam: PF00155; aminotran\_1.2; I.  
DR PRINTS: PR00753; ACCSYNTHASE.  
DR PROSITE: PS01048; RIBOSOMAL\_S6; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 432 AA; 49429 MW; 14D1602E32DB31FE CRC64;

Query Match 78.0%; Score 39; DB 17; Length 432;  
Best Local Similarity 77.8%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CSVPLTSVC 9  
Db 389 CWPLTSPC 397

RESULT 14  
Q8WVW3 PRELIMINARY; PRT; 323 AA.  
AC Q8WVW3;  
DT 01-MAR-2002 (TRENBLrel. 20, Created)  
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Strausberg R.;  
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC017468; AAH17468.1; -  
DR InterPro: IPR008928; Glyco\_trans\_6hp.  
KW Hypothetical protein.  
FT NON TER 1  
SQ SEQUENCE 323 AA; 35965 MW; BCC0D79E4A13CCE3 CRC64;

Query Match 76.0%; Score 38; DB 4; Length 323;  
Best Local Similarity 66.7%; Pred. No. 16;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CSVPLTSVC 9  
Db 307 CSVPTDPC 315

RESULT 15  
Q9H6R1 PRELIMINARY; PRT; 491 AA.  
AC Q9H6R1;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ21969.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

Tanaka T., Nakamura Y., Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AK025622; BAB15192.1; -  
DR InterPro: IPR008928; Glyco\_trans\_6hp.  
KW Hypothetical protein.  
SQ SEQUENCE 491 AA; 54720 MW; 4F73721A9D092C37 CRC64;  
Query Match 76.0%; Score 38; DB 4; Length 491;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CSVPLTSVC 9  
Db 475 CSVPTDPC 483

Search completed: September 5, 2004, 10:00:01  
Job time : 22.3636 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:37:49 ; Search time 29.5455 Seconds  
(without alignments)  
86.068 Million cell updates/sec

Title: US-09-761-636A-10

Perfect score: 50

Sequence: 1 CSVPLTSCV 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Genesep\_29Jan04:.\*  
1: Genesep1980s:.\*  
2: Genesep1990s:.\*  
3: Genesep2000s:.\*  
4: Genesep2001s:.\*  
5: Genesep2002s:.\*  
6: Genesep2003as:.\*  
7: Genesep2003bs:.\*  
8: Genesep2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	9	AAU04529	Aau04529 VEGF base
2	43	86.0	1487	2 AAB23826	Aab23826 Murine ph
3	39	78.0	73	4 AAU63687	Aau63687 Propionib
4	39	78.0	73	6 ABM60206	Abm60206 Propionib
5	39	78.0	88	4 AAU53277	Aau53277 Propionib
6	39	78.0	88	6 ABM49796	Abm49796 Propionib
7	38	76.0	101	4 AAU43894	Aau43894 Propionib
8	38	76.0	101	6 ABM40413	Abm40413 Propionib
9	38	76.0	121	5 AAU74551	Aau74551 Human pro
10	38	76.0	225	4 ABG11475	Abg11475 Novel hum
11	38	76.0	742	4 AAB74603	Aab74603 Human pro
12	38	76.0	802	4 AAE01782	Aae01782 Human gen
13	38	76.0	802	5 ABG64141	Abg64141 Human alb
14	38	76.0	802	6 ADA56848	Ada56848 Human sec
15	38	76.0	802	6 ADA40695	Ada40695 Human sec
16	38	76.0	802	6 ABR47707	Abr47707 Human sec
17	38	76.0	802	7 ADC74086	Adc74086 Human sec
18	38	76.0	802	7 ADD37860	Add37860 Human sec
19	37	74.0	9	5 ABJ04424	Abj04424 Stem cell
20	37	74.0	86	4 AAU58469	Aau58469 Propionib
21	37	74.0	86	6 ABM54988	Abm54988 Propionib
22	37	74.0	1929	4 ABB67846	Abb67846 Drosophil
23	36	72.0	53	5 ABP06465	Abp06465 Human ORF
24	36	72.0	90	4 AAG74036	Aag74036 Human col
25	36	72.0	133	4 AAU21569	Aau21569 Novel hum

26	36	72.0	133	4 AAU19674	Aau19674 Human nov
27	36	72.0	133	4 AAM43498	Aam43498 Human pol
28	36	72.0	133	4 ABB10143	Abb10143 Human GDN
29	36	72.0	133	4 AAU18408	Aau18408 Human end
30	36	72.0	133	5 ABP47894	Abp47894 Human pol
31	36	72.0	133	5 ABP66730	Abp66730 Human pol
32	36	72.0	133	7 ADC10856	Adc10856 Human ext
33	36	72.0	133	7 ADC46210	Adc46210 Human neo
34	36	72.0	201	2 AAWS2837	Aaw52837 Human C-t
35	36	72.0	201	2 AAY27449	Aay27449 Human SDC
36	36	72.0	201	2 AAY41764	Aay41764 Human PRO
37	36	72.0	201	2 AAW73888	Aaw73888 Human DC3
38	36	72.0	201	3 AAB44320	Aab44320 Human PRO
39	36	72.0	201	4 AAU29077	Aau29077 Human PRO
40	36	72.0	201	5 ABB90372	Abb90372 Human pol
41	36	72.0	201	6 ABU58453	Abu58453 Human PRO
42	36	72.0	201	6 ABU88001	Abu88001 Novel hum
43	36	72.0	201	6 ABU84316	Abu84316 Human sec
44	36	72.0	201	6 ABR66190	Abr66190 Human sec
45	36	72.0	201	6 ABR65580	Abr65580 Human sec

## ALIGNMENTS

RESULT 1  
AAU04529  
ID AAU04529 standard; peptide; 9 AA.  
XX AC AAU04529;  
DT 26-SEP-2001 (first entry)  
XX DE VEGF based monocyclic peptide 7.  
XX KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
KW diabetic retinopathy; chronic inflammation; cyclic.  
XX OS Synthetic.  
XX FH Key  
FH Disulfide-bond 1..9  
FH Location/Qualifiers  
FT /note= "This bond cyclises the peptide"  
XX PN WO200152875-A1.  
XX PD 26-JUL-2001.  
XX PF 18-JAN-2001; 2001WO-US001533.  
XX PR 18-JAN-2000; 2000US-0176293P.  
XX PR 16-MAY-2000; 2000US-0204590P.  
XX PA (LUDW-) LUDWIG INST CANCER RES.  
XX PI Achen MG, Hughes RA, Stacker S, Cendron A;  
XX WPI; 2001-442248/47.  
XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
from an exposed loop of a growth factor protein by oxidizing the cysteine  
residues.  
XX Claim 49; Page 32; 102pp; English.  
XX The sequence represents a monomeric monocyclic peptide of the invention,  
whose 3-dimensional structure is modelled on the expose loop of human  
CC VEGPD (vascular endothelial growth factor). The invention relates to a  
method of producing a monomeric monocyclic peptide by a measuring beta-  
beta carbon separation distances on opposite antiparallel strands of a

peptide loop fragment from an exposed loop of a growth factor protein and cyclising the peptide by oxidising the cysteine residues. The monocyclic peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic peptides) and a cyclic peptide with at least one amino acid deleted prior to cyclisation are used to interfere with angiogenesis, neovascularisation or lymphangiogenesis in a mammal with a condition characterised by angiogenesis, neovascularisation or lymphangiogenesis. The condition is diabetic retinopathy, psoriasis, arthropathy, hemangioma, vascularised malignant or benign tumour, post-recovery cerebrovascular accident, post-angioplasty restenosis, head, heat or cold trauma, substance-induced neovascularisation of the liver, excessive hormone-related angiogenic dysfunction, diabetes induced neovascular sequelae, hypertension induced neovascular sequelae, or chronic liver infection. The peptides are also used to modulate vascular permeability in a mammal (the mammal has a condition characterised by fluid accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura, or brain. The peptides are used to image blood vessels and lymphatic vasculature. The monomeric and bicyclic peptides are used to interfere with at least one biological activity induced by VEGF, VEGF-C or -D and are also used in combination with an anti-inflammatory agent, to treat a chronic inflammation, especially rheumatoid arthritis, psoriasis and diabetic retinopathy

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 50; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
| | | | | | | |  
Db 1 CSVPLTSVC 9

RESULT 2

AAB23826  
ID AAB23826 standard; protein; 1487 AA.

XX AAB23826;

XX 15-JAN-2001 (first entry)

XX Murine phospholipase A2 receptor protein SEQ ID NO:7.

DE Endocytic C lectin family; E-selectin; type C lectin; identification.

XX Mus sp.

XX US6117977-A.

XX 12-SEP-2000.

XX 24-APR-1997; 97US-00840062.

XX 24-APR-1996; 96US-0052524P.

XX 24-APR-1996; 96US-00637021.

XX (GETH ) GENENTECH INC.

XX Wu K, Lasky LA;

XX WPI; 1997-535838/49.

XX Human and mouse type C lectin(s) - useful as competitive inhibitor of lectin activity and as molecular markers for tissues that express them.

XX Example; Fig 3; 72pp; English.

XX The present invention describes an isolated type C lectin polypeptide (I) comprising amino acid residues 37-1393, 37-174, 175-229, 234-360, 381-507, 520-645, 667-809, 824-951, 970-1108, 1110-1243, or 1259-1393 of the protein sequence given in AAB23822. The first 2 polypeptides are capable of binding to a carbohydrate residue, and the rest of the polypeptides

are useful for producing antibodies capable of binding to these 2 polypeptides. Polynucleotide sequence encoding the polypeptides of the invention are useful in the identification and purification of their native ligands, and as molecular markers of the tissues in which they are expressed. They provide valuable sequence motifs, which can be inserted or substituted into other native members of the endocytic type C lectins, and provide hybridisation probes for searching cDNA and genomic libraries for the coding sequence of other type C lectins. Variants of type C lectins may be used therapeutically as competitive inhibitors of the biological activity of native type C lectins. The present sequence represents a murine phospholipase A2 receptor protein which is homologous to a type C lectin

XX SQ Sequence 1487 AA;

Query Match 86.0%; Score 43; DB 2; Length 1487;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
| | | | | | | |  
Db 929 CSVPLPSIC 937

RESULT 3

AAU63687  
ID AAU63687 standard; protein; 73 AA.

XX AAU63687;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #24583.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis; uveitis; endophthalmitis; bone joint; central nervous system; ELISA; inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay; dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59634.

XX Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris.

XX Example 1; SEQ ID NO 24882; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the



xx The invention relates to an isolated polynucleotide (ACF644435-ACF64733) encoding a *Propionibacterium* acnes protein. The invention also relates to CC polypeptides encoded by the polynucleotides (ABM56244-ABM64536) and to CC immunogenic fragments of *P. acnes* polypeptides. The invention CC additionally encompasses expression vectors and host cells comprising CC a polynucleotide of the invention; antibodies against polypeptides of the CC polynucleotide of the invention; a polypeptide of the invention; a CC invention; fusion proteins comprising a polypeptide of the invention; a CC

XX (CORI-) CORIXA CORP.  
PA  
XX  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
XX WPI: 2001-616774/71.  
DR  
DR N-PSDB; AAS59521.  
XX  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.  
XX  
XX  
PS Example 1; SEQ ID NO 5089; 1069pp; English.  
XX

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA). Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 101 AA;

Query Match 76.0%; Score 38; DB 4; Length 101;  
Best Local Similarity 66.7%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPPLTSVC 9  
|::||| ||  
Db 16 CNLPLTLVC 24

RESULT 8  
ABM40413  
ID ABM40413 standard; protein; 101 AA.  
XX  
AC ABM40413;

20-OCT-2003 (first entry)

Propionibacterium acnes predicted ORF-encoded polypeptide #5089.

Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
immunostimulant; immune response; vaccine.

Propionibacterium acnes.

WO2003033515-A1.

24-APR-2003.

11-OCT-2002; 2002WO-US032727.

15-OCT-2001; 2001US-00978825.

(CORI-) CORIXA CORP.

Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
Barth B, Valliave-Douglass J;

WPI: 2003-381789/36.  
N-PSDB; ACP64450.

New Propionibacterium acnes polypeptides and polynucleotides encoding the polypeptide, useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein.

Example 1; SEQ ID NO 5089; 1481pp; English.

The invention relates to an isolated polynucleotide (ACF64435-ACF64733) encoding a Propionibacterium acnes protein. The invention also relates to polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to

CC immunogenic fragments of P. acnes polypeptides. The invention additionally encompasses expression vectors and host cells comprising a polynucleotide of the invention; antibodies against polypeptides of the invention; fusion proteins comprising a polypeptide of the invention; a method for stimulating an immune response specific for a P. acnes polypeptide and an isolated T cell population comprising T cells prepared via this method; a vaccine composition (comprising P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations, or antigen-presenting cells that express the polypeptide); a method and kit for detecting or determining the presence or absence of P. acnes in a patient; and a method for inhibiting the development of P. acnes in a patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations or antigen-presenting cells that express the polypeptides are useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein. The polynucleotides can also be used as probes or primers for nucleic acid hybridisation. The vaccine composition is useful for the stimulation of an immune response against P. acnes, or for treating acne, and the kit is useful for performing a diagnostic assay. The present sequence represents a polypeptide predicted to be encoded by an ORF (open reading frame) contained within the P. acnes polynucleotides of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 101 AA;

Query Match 76.0%; Score 38; DB 6; Length 101;  
Best Local Similarity 66.7%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPPLTSVC 9  
|::||| ||  
Db 16 CNLPLTLVC 24

RESULT 9  
AAU74551

ID AAU74551 standard; protein; 121 AA.

XX  
AC AAU74551;

08-MAY-2002 (first entry)

Human protein phosphatase 13.31 polypeptide.

Human; protein phosphatase 13.31; enzyme; malignant tumour; cancer;  
embryonic developmental abnormality; autoimmune disease; antisense;  
gynaecological; cytostatic; immunosuppressant; gene therapy.

Homo sapiens.

WO200212457-A1.

14-FEB-2002.

02-JUL-2001; 2001WO-CN001118.

07-JUL-2000; 2000CN-00117017.

(BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.

Mao Y, Xie Y;

WPI: 2002-172153/22.  
N-PSDB; ABK14072.

Human protein phosphatase 13.31 and encoding polynucleotide, used in diagnosis and treatment of malignant tumors, autoimmune disease and embryonic development abnormality.

Claim 1; Page 30; 34pp; Chinese.

CC The invention relates to an isolated polypeptide of human protein  
 CC phosphatase 13.31 and its associated polynucleotide. The sequences are  
 CC used in diagnosis and treatment of malignant tumour, embryonic  
 CC developmental abnormality and autoimmune disease and for studying human  
 CC antisense. This sequence represents human protein phosphatase 13.31  
 XX  
 SQ Sequence 121 AA;  
 Query Match 76.0%; Score 38; DB 5; Length 121;  
 Best Local Similarity 66.7%; Pred. No. 78;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CSVPLTSVC 9  
 |||||:  
 DB 105 CSVBITDPC 113  
 RESULT 10  
 ABG11475  
 ID ABG11475 standard; protein; 225 AA.  
 AC  
 AC ABG11475;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #11466.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW  
 OS Homo sapiens.  
 XX  
 XX WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS75662.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PT  
 XX Claim 20; SEQ ID NO 41834; 103pp; English.  
 PS  
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC amino acid sequences of the invention.

CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 225 AA;  
 Query Match 76.0%; Score 38; DB 4; Length 225;  
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CSVPLTSVC 9  
 |||||:  
 DB 209 CSVBITDPC 217  
 RESULT 11  
 AAB74693  
 ID AAB74693 standard; protein; 742 AA.  
 XX  
 AC AAB74693;  
 XX  
 DT 12-JUN-2001 (first entry)  
 XX  
 DE Human protease and protease inhibitor PPIM-26.  
 XX  
 KW Human; protease; protease inhibitor; protease and protease inhibitor;  
 KW PPIIM; identification; diagnosis; anti-human immunodeficiency virus; HIV;  
 KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;  
 KW antithyroid; immunosuppressive; nephrotropic; antigout; thyromimetic;  
 KW cytostatic; antibacterial; fungicide; protozoicide; antiarteriosclerotic;  
 KW antiatherosclerotic; antipsoriatic; virucide; hepatotropic; gene therapy;  
 KW autoimmune disorder; inflammatory disease; SCID; Chediak-Higashi syndrome;  
 KW severe combined immunodeficiency disease; AIDS; Dideozyme's syndrome;  
 KW Cushing's disease; Addison's disease; autoimmunity; thyroiditis; gout;  
 KW Crohn's disease; diabetes mellitus; Good pasture's syndrome; infection;  
 KW Grave's disease; Hashimoto's thyroiditis; Sjogren's syndrome; cancer;  
 KW Werner's syndrome; cell proliferative disorder; arteriosclerosis;  
 KW atherosclerosis; cirrhosis; hepatitis; psoriasis.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200110903-A2.  
 XX  
 PD 15-FEB-2001.  
 XX  
 XX 09-AUG-2000; 2000WO-US021878.  
 XX  
 XX 09-AUG-1999; 99US-0147986P.  
 PR 21-OCT-1999; 99US-0160807P.  
 XX  
 XX (INCV-) INCVTE GENOMICS INC.  
 PA  
 XX Yue H, Lal P, Tang YT, Bandman O, Baughn MR, Azimzai Y, Lu DAM;  
 PI Yang J;  
 XX  
 XX WPI; 2001-202760/20.  
 DR N-PSDB; AAF81739.  
 DR  
 XX New protease (inhibitors) useful for diagnosis and treatment of  
 PT autoimmune/inflammatory disorders such as acquired immunodeficiency  
 PT syndrome, Cushing's disease, Addison's disease and cell proliferative  
 PT disorders such as cancer.  
 PT  
 XX Claim 1; Page 114-115; 134pp; English.  
 PS  
 XX AAF81714 to AAF81740 encode the human proteases and protease inhibitors  
 CC (PPIIMs) given in AAB74668 to AAB74694. The PPIIMs can have activities such  
 CC as: anti-human immunodeficiency virus (HIV); antidiabetic; antithyroid;  
 CC immunostimulant; immunomodulator; antiinflammatory; immunosuppressive;  
 CC nephrotropic; antigout; thyromimetic; cytostatic; antibacterial;  
 CC fungicide; protozoicide; antiarteriosclerotic; antiatherosclerotic;  
 CC virucide; antipsoriatic; and hepatotropic. PPIIM polynucleotide and  
 CC protein sequences can be used in the diagnosis, treatment and prevention

CC of autoimmune/inflammatory disorders such as AIDS, DiGeorge's syndrome,  
 CC severe combined immunodeficiency disease (SCID), Chediak-Higashi  
 CC syndrome, Cushing's disease, Addison's disease, autoimmune thyroiditis,  
 CC Crohn's disease, diabetes mellitus, Good pasture's syndrome, gout,  
 CC Grave's disease, Hashimoto's thyroiditis, Sjogren's syndrome, Werner's  
 CC syndrome, viral, bacterial, fungal, parasitic, protozoal, and helminthic  
 CC infections and cell proliferative disorder such as arteriosclerosis,  
 CC atherosclerosis, cirrhosis, hepatitis, psoriasis and cancer. PPM  
 CC polynucleotide sequences can be used in somatic or germline gene therapy  
 CC and in diagnosis of diseases. They can also be used in generating  
 CC hybridisation probes useful in mapping the naturally occurring genomic  
 CC sequences and in molecular biology techniques  
 XX  
 SQ Sequence 742 AA;

Query Match 76.0%; Score 38; DB 4; Length 742;  
 Best Local Similarity 66.7%; Pred. No. 4.6e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPPLTSVC 9  
 Db 726 CSVPITDPC 734  
 |||||:|

## RESULT 12

AAE01782  
 ID AAE01782 standard; protein; 802 AA.

AC AAE01782;

DT 17-JUL-2001 (first entry)

XX Human gene 13 encoded secreted protein HDPNW93, SEQ ID NO:103.

XX Human; secreted protein; proliferative disorder; cancer; tumour;

KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; chromosome 17; binding partner identification.  
 OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Peptide 1..19

FT Protein /label= Signal\_peptide

FT 20..802

FT /label= Human\_mature\_secreted\_protein

XX WO200134627-A1.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000WO-US030628.

XX 12-NOV-1999; 99US-0164744P.

XX 30-JUN-2000; 2000US-0215140P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;

XX WPI; 2001-316491/33.

XX N-PSDB; AAD05591.

XX New nucleic acid molecules encoding human secreted proteins, used in  
 PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and  
 PT Parkinson's diseases and cancers.  
 XX

PS Claim 11; Page 492-494; 567pp; English.

XX AAD05579-AAD05658 represent cDNAs corresponding to 28 human secreted  
 CC protein genes and AAE01770-AAE01849 represent the proteins they encode.  
 CC AAE01850-AAE01860 represent secreted human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 28 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein of the invention  
 XX  
 SQ Sequence 802 AA;

Query Match 76.0%; Score 38; DB 4; Length 802;  
 Best Local Similarity 66.7%; Pred. No. 5e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPPLTSVC 9  
 |||||:|

Db 786 CSVPITDPC 794

## RESULT 13

ABG64141

ID ABG64141 standard; protein; 802 AA.

XX AC ABG64141;

XX DT 27-AUG-2002 (first entry)

XX DE Human albumin fusion protein #816.

XX KW Albumin fusion protein; therapeutic protein X; human albumin; HA;

XX KW human serum albumin; HSA; cancer; reproductive disorder;

XX KW digestive disorder; immune disorder; endocrine disorder;

XX KW haematopoietic disorder; neural disorder; connective disorder;

XX KW cytostatic; antiinfertility; antiinflammatory; antitumor;

XX KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neuroleptic;

XX KW osteopathic; antiarthritic.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200177137-A1.

XX PD 18-OCT-2001.

XX PF 12-APR-2001; 2001WO-US011988.

XX PR 12-APR-2000; 2000US-0229358P.

XX PR 25-APR-2000; 2000US-0199384P.

XX PR 21-DEC-2000; 2000US-0256931P.

XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Haseltine WA;  
XX WPI; 2002-010886/01.  
XX New fusion protein for treating disease e.g. diabetes comprises an  
XX albumin fused to a therapeutic protein.  
XX Claim 1; Page 1035-1037; 2102pp; English.  
XX The present invention relates to albumin fusion proteins comprising a  
XX therapeutic protein X and human albumin (HA), also known as human serum  
XX albumin, HSA). The proteins are useful for treating a disease or disorder  
XX that may be modulated by therapeutic protein X. The albumin extends the  
XX shelf-life of protein X, and may increase its biological in vitro/in vivo  
XX activity. The protein is useful for treating and diagnosing disorders  
XX such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
XX disease, ulcerative colitis), immune disorders (e.g. acquired  
XX immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
XX haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
XX Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
XX schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
XX ABG63326-ABG65518 represent albumin fusion proteins of the invention  
XX SQ Sequence 802 AA;  
XX Query Match 76.0%; Score 38; DB 5; Length 802;  
XX Best Local Similarity 66.7%; Pred. No. 5e+02;  
XX Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
XX QY 1 CSVPITDPC 9  
XX Db 786 CSVPITDPC 794  
XX RESULT 14  
XX ADA56848  
XX ID ADA56848 standard; protein; 802 AA.  
XX AC ADA56848;  
XX XX 20-NOV-2003 (first entry)  
XX DT 20-NOV-2003 (first entry)  
XX DE Human secreted protein #131.  
XX KW immunosuppressive; anti-inflammatory; antiaesthetic; antiallergic;  
XX KW cytotactic; cerebroprotective; neuroprotective; nootropic;  
XX KW cardiovascular; antiatherosclerotic; gene therapy;  
XX KW human secreted protein; immune disorder; inflammation;  
XX KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;  
XX KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;  
XX KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;  
XX KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;  
XX KW triple helix formation; antisense gene therapy; forensic biology.  
XX OS Homo sapiens.  
XX XX WO2002102994-A2.  
XX PN 27-DEC-2002.  
XX PD 19-MAR-2002; 2002WO-US008278.  
XX XX 21-MAR-2001; 2001US-0277340P.  
XX PR 19-JUL-2001; 2001US-0306171P.  
XX PR 13-NOV-2001; 2001US-0331287P.  
XX XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX PI 27-DEC-2002.  
XX PD

DR WPI; 2003-167512/16.  
DR N-PSDB; ADA55952.  
XX New human secreted polypeptides and polynucleotides, useful for  
XX diagnosing, treating or preventing e.g. immune disorders, inflammatory  
XX conditions, respiratory disorders, cancers, CNS disorders, or  
XX neurodegenerative disorders.  
XX Claim 13; SEQ ID NO 1038; 1754pp; English.  
XX The invention relates to 592 new human secreted polypeptides useful for  
XX diagnosing, treating or preventing e.g. immune disorders, inflammatory  
XX conditions, respiratory disorders, cancers, CNS disorders, or  
XX neurodegenerative disorders, or polypeptides comprising an amino acid  
XX sequence at least 95% identical to the new sequences. The polypeptides,  
XX antibodies or antibody fragments that bind to the polypeptides, nucleic  
XX acids encoding the polypeptides, agonists or antagonists that binds to  
XX the polypeptide, are useful in preparing diagnostic or pharmaceutical  
XX compositions for diagnosing, treating or preventing an e.g. immune  
XX disorders, inflammatory conditions (e.g. inflammatory bowel disease,  
XX nephritis or Crohn's disease), respiratory disorders (e.g. asthma and  
XX allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders  
XX (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative  
XX disorders (e.g. Parkinson's disease or Alzheimer's disease), and  
XX cardiovascular disorders (e.g. atherosclerosis or myocarditis). The  
XX polynucleotides are useful for chromosome identification, chromosome  
XX mapping, for controlling gene expression through triple helix formation  
XX or antisense DNA or RNA, in gene therapy, for identifying individuals  
XX from minute biological samples, in forensic biology, and as hybridization  
XX probes. The polypeptides are useful for as molecular weight markers on  
XX sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)  
XX gels, to raise antibodies, for testing biological activities, and for  
XX treating or preventing neural disorders, immune system disorders,  
XX muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
XX renal, proliferative and/or cancerous diseases. This sequence corresponds  
XX to one of the polypeptide of the invention. Note: The sequence data for  
XX this patent did form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 802 AA;  
XX Query Match 76.0%; Score 38; DB 6; Length 802;  
XX Best Local Similarity 66.7%; Pred. No. 5e+02;  
XX Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
XX QY 1 CSVPITDPC 9  
XX Db 786 CSVPITDPC 794  
XX RESULT 15  
XX ADA40695  
XX ID ADA40695 standard; protein; 802 AA.  
XX AC ADA40695;  
XX XX 20-NOV-2003 (first entry)  
XX DT 20-NOV-2003 (first entry)  
XX DE Human secreted protein.  
XX KW Human; secreted protein; cancer; hyperproliferative disorder;  
XX KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;  
XX KW anaemia; allergic reaction; asthma; cardiovascular disorder;  
XX KW wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;  
XX KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;  
XX KW vulnery; cardiant; gene therapy.  
XX OS Homo sapiens.  
XX XX WO2002102993-A2.  
XX PN 27-DEC-2002.  
XX PD



```

XX 19-MAR-2002; 2002WO-US008123.
PF
XX
XX 21-MAR-2001; 2001US-0277340P.
PR
XX 19-JUL-2001; 2001US-0306171P.
PR
XX 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX WPI; 2003-175238/17.
DR
XX
XX New human secreted proteins and nucleic acid molecules, useful for
PT preparing a diagnostic or pharmaceutical composition for diagnosing,
PT preventing or treating cancer or other hyperproliferative disorder,
PT asthma, allergies or AIDS.
XX
XX Claim 1; SEQ ID NO 1077; 3205pp; English.
PS
XX
XX The invention relates to novel genes ADA39629-ADA40565 and proteins
CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The polypeptides, nucleic acid molecules, antibodies or their
CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 802 AA;

Query Match 76.0%; Score 38; DB 6; Length 802;
Best Local Similarity 56.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPLTVC 9
Db 786 CSVPTDPC 794

Search completed: September 5, 2004, 09:55:14
Job time : 32.5455 secs

```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 10:00:15 ; Search time 25.2727 Seconds  
(without alignments)  
112.199 Million cell updates/sec

Title: US-09-761-636A-10  
Perfect score: 50  
Sequence: 1 CSVPLTSVC 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	9	9	US-09-761-636A-10
2	42	84.0	154	12	US-10-424-599-241613
3	42	84.0	194	12	US-10-424-599-241616
4	41	82.0	78	12	US-10-424-599-206581
5	41	82.0	359	12	US-10-424-599-174316
6	41	82.0	549	12	US-10-424-599-205890
7	41	82.0	565	12	US-10-425-114-69391
8	41	82.0	883	16	US-10-437-963-133870
9	39	78.0	63	12	US-10-424-599-276101
10	39	78.0	432	15	US-10-369-493-1081
11	38	76.0	802	11	US-09-833-245-888
12	37	74.0	82	12	US-10-424-599-169061
13	37	74.0	95	12	US-10-424-599-276082
14	37	74.0	167	12	US-10-424-599-243236
15	37	74.0	215	12	US-10-425-114-56269

16	37	74.0	267	12	US-10-424-599-236212	Sequence 236212, App
17	37	74.0	269	12	US-10-425-114-39180	Sequence 39180, A
18	37	74.0	718	16	US-10-437-963-160508	Sequence 160508, App
19	36	72.0	90	14	US-10-106-698-4810	Sequence 4810, App
20	36	72.0	133	9	US-09-764-870-324	Sequence 324, App
21	36	72.0	133	9	US-09-764-853-451	Sequence 451, App
22	36	72.0	133	14	US-10-125-540-324	Sequence 324, App
23	36	72.0	133	14	US-10-103-313-296	Sequence 296, App
24	36	72.0	133	15	US-10-158-057-176	Sequence 176, App
25	36	72.0	201	9	US-09-978-295A-477	Sequence 477, App
26	36	72.0	201	9	US-09-978-697-477	Sequence 477, App
27	36	72.0	201	9	US-09-978-192A-477	Sequence 477, App
28	36	72.0	201	9	US-09-999-832A-477	Sequence 477, App
29	36	72.0	201	10	US-09-978-189-477	Sequence 477, App
30	36	72.0	201	10	US-09-978-608A-477	Sequence 477, App
31	36	72.0	201	10	US-09-978-585A-477	Sequence 477, App
32	36	72.0	201	10	US-09-978-191A-477	Sequence 477, App
33	36	72.0	201	10	US-09-978-403A-477	Sequence 477, App
34	36	72.0	201	10	US-09-978-564A-477	Sequence 477, App
35	36	72.0	201	10	US-09-999-833A-477	Sequence 477, App
36	36	72.0	201	10	US-09-981-915A-477	Sequence 477, App
37	36	72.0	201	10	US-09-978-824-477	Sequence 477, App
38	36	72.0	201	10	US-09-918-585A-477	Sequence 477, App
39	36	72.0	201	10	US-09-978-423A-477	Sequence 477, App
40	36	72.0	201	10	US-09-999-830A-477	Sequence 477, App
41	36	72.0	201	10	US-09-978-757A-477	Sequence 477, App
42	36	72.0	201	10	US-09-978-187B-477	Sequence 477, App
43	36	72.0	201	10	US-09-978-643A-477	Sequence 477, App
44	36	72.0	201	10	US-09-978-375A-477	Sequence 477, App
45	36	72.0	201	10	US-09-978-375A-477	Sequence 477, App

## ALIGNMENTS

RESULT 1  
US-09-761-636A-10  
; Sequence 10, Application US/09761636A  
; Patent No. US20020065218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09/761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-761-636A-10

Query Match 100.0%; Score 50; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
| | | | | | | | |  
Db 1 CSVPLTSVC 9

RESULT 2  
US-10-424-599-241613  
; Sequence 241613, Application US/10424599  
; Publication No. US20040031072A1

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241613
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60203C.1.pep
US-10-424-599-241613

Query Match      84.0%; Score 42; DB 12; Length 154;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 CSVPLTSVC 9
Db 22 CSSPITSVC 30

RESULT 3
US-10-424-599-241616
; Sequence 241616, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241616
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60206C.1.pep
US-10-424-599-241616

Query Match      84.0%; Score 42; DB 12; Length 194;
Best Local Similarity 77.8%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9
Db 66 CSSPITSVC 74

RESULT 4
US-10-424-599-206581
; Sequence 206581, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 206581
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_28570C.1.pep
US-10-424-599-206581

Query Match      82.0%; Score 41; DB 12; Length 78;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9
Db 57 CSSPITSVC 65

RESULT 5
US-10-424-599-174316
; Sequence 174316, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 174316
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: unsure
; LOCATION: (1)..(359)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_128426C.1.pep
US-10-424-599-174316

Query Match      82.0%; Score 41; DB 12; Length 359;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 8
Db 151 CSVPLTSVC 158

RESULT 6
US-10-424-599-205890
; Sequence 205890, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 205890
; LENGTH: 549
; TYPE: PRT
```

```
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_27947C.1.pep
US-10-424-599-205990

Query Match      82.0%; Score 41; DB 12; Length 549;
Best Local Similarity 66.7%; Pred. No. 82;
Matches      6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CSVPLTSVC 9
Db      66 CSSPITSIC 74

RESULT 7
US-10-425-114-69391
; Sequence 69391, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53113)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69391
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMROPIC11B04_FLI.pep
US-10-425-114-69391

Query Match      82.0%; Score 41; DB 12; Length 565;
Best Local Similarity 66.7%; Pred. No. 85;
Matches      6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 CSVPLTSVC 9
Db      82 CSSPITSIC 90

RESULT 8
US-10-437-963-133870
; Sequence 133870, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 133870
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3569C.1.pep
```

US-10-437-963-133870

```
Query Match      82.0%; Score 41; DB 16; Length 883;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 CSVPLTSVC 8
Db      655 CSVPLTSVC 662
```

RESULT 9

```
US-10-424-599-276101
; Sequence 276101, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 276101
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_91339C.1.pep
US-10-424-599-276101
```

```
Query Match      78.0%; Score 39; DB 12; Length 63;
Best Local Similarity 66.7%; Pred. No. 23;
Matches      6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 CSVPLTSVC 9
Db      9 CSSPLSSIC 17
```

RESULT 10

```
US-10-369-493-1081
; Sequence 1081, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1081
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-10-369-493-1081
```

```
Query Match      78.0%; Score 39; DB 15; Length 432;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches      7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 CSVPLTSVC 9
Db      11 CSSPLSSIC 17
```

Db 389 CVPPLTSFC 397

## RESULT 11

US-09-833-245-888  
; Sequence 888, Application US/09833245  
; Publication No. US20040010134A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229, 358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256, 931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199, 384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 888  
; LENGTH: 802  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-245-888

Query Match 76.0%; Score 38; DB 11; Length 802;  
Best Local Similarity 66.7%; Pred. No. 3.9e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPPLTSVC 9  
Db 786 CVPPLTDC 794

## RESULT 12

US-10-424-599-169061  
; Sequence 169061, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 169061  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_123677C.1.pap  
US-10-424-599-169061

Query Match 74.0%; Score 37; DB 12; Length 82;  
Best Local Similarity 66.7%; Pred. No. 64;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPPLTSVC 9  
Db 50 CVPPLTATC 58

## RESULT 13

US-10-424-599-276082  
; Sequence 276082, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 276082  
; LENGTH: 95  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(95)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_91321C.1.pap  
US-10-424-599-276082

Query Match 74.0%; Score 37; DB 12; Length 95;  
Best Local Similarity 66.7%; Pred. No. 73;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPPLTSVC 9  
Db 1 CVPPLTGLC 9

## RESULT 14

US-10-424-599-243236  
; Sequence 243236, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 243236  
; LENGTH: 167  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_61670C.1.pap  
US-10-424-599-243236

Query Match 74.0%; Score 37; DB 12; Length 167;  
Best Local Similarity 55.6%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CVPPLTSVC 9  
Db 107 CVPPLTKIC 115

## RESULT 15

US-10-425-114-56269  
; Sequence 56269, Application US/10425114  
; Publication No. US20040034868A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 56269  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700835470\_FLI.pep  
US-10-425-114-56269

Query Match 74.0%; Score 37; DB 12; Length 215;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CSVPPLTATC 9  
Db 183 CSVPPLTATC 191

Search completed: September 5, 2004, 10:29:20  
Job time : 26.2727 secs

**This Page Blank (uspto)**

**ge Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:55:30 ; Search time 8 seconds  
(without alignments)  
58.079 Million cell updates/sec

Title: US-09-761-636A-10  
Perfect score: 50  
Sequence: 1 CSVPLTSVC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	86.0	1487	3	US-08-840-062-7
2	41	82.0	422	4	US-09-489-039A-10842
3	36	72.0	168	4	Sequence 10842, A
4	36	72.0	201	2	US-09-252-991A-20364
5	36	72.0	201	2	US-08-688-342-1
6	35	70.0	359	3	US-09-113-788-1
7	34	68.0	134	4	US-08-586-165-4
8	34	68.0	386	1	US-09-621-976-4725
9	34	68.0	386	1	US-08-134-012-3
10	34	68.0	386	1	US-08-520-519-3
11	34	68.0	400	4	US-09-039-798-3
12	34	68.0	747	2	US-09-673-395A-479
13	34	68.0	747	3	US-08-816-693A-51
14	34	68.0	747	3	US-08-885-291-51
15	33	66.0	126	4	US-09-496-672-51
16	33	66.0	254	4	US-09-402-016A-4
17	33	66.0	403	4	US-09-543-681A-4919
18	32	64.0	109	4	US-09-543-681A-6083
19	32	64.0	109	4	US-09-469-186-1
20	32	64.0	134	3	US-09-469-186-1
21	32	64.0	171	4	US-08-543-246B-20
22	32	64.0	197	4	US-09-252-991A-26927
23	32	64.0	216	3	US-09-431-888-8
24	32	64.0	216	3	US-08-543-246B-9
25	32	64.0	228	4	US-08-543-246B-24
26	32	64.0	276	4	US-09-489-039A-13837
27	32	64.0	294	4	US-09-252-991A-18128
					Sequence 30882, A

28 32 64.0 321 3 US-08-915-795-9 Sequence 9, Appli  
29 32 64.0 325 3 US-08-915-795-3 Sequence 3, Appli  
30 32 64.0 353 4 US-09-484-970B-171 Sequence 171, Appl  
31 32 64.0 354 3 US-08-915-795-5 Sequence 5, Appli  
32 32 64.0 358 3 US-08-915-795-8 Sequence 8, Appli  
33 32 64.0 379 2 US-07-857-224B-87 Sequence 87, Appli  
34 32 64.0 497 4 US-09-134-000C-5990 Sequence 5990, Ap  
35 32 64.0 505 1 US-08-220-603A-10 Sequence 10, Appl  
36 32 64.0 1107 4 US-09-358-383C-16 Sequence 16, Appl  
37 32 64.0 1463 1 US-08-220-603A-11 Sequence 11, Appl  
38 31 62.0 14 1 US-08-297-633A-5 Sequence 5, Appli  
39 31 62.0 14 2 US-08-485-721-5 Sequence 5, Appli  
40 31 62.0 14 2 US-08-392-935-5 Sequence 5, Appli  
41 31 62.0 14 3 US-08-897-236-5 Sequence 5, Appli  
42 31 62.0 14 3 US-09-167-874-5 Sequence 5, Appli  
43 31 62.0 14 4 US-09-500-253B-5 Sequence 5, Appli  
44 31 62.0 14 5 PCT-US93-08325-5 Sequence 5, Appli  
45 31 62.0 14 5 PCT-US93-08326-5 Sequence 5, Appli

## ALIGNMENTS

RESULT 1  
US-08-840-062-7  
; Sequence 7, Application US/08840062  
; Patent No. 6117977  
; GENERAL INFORMATION:  
; APPLICANT: LASKY, LAURENCE A.  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: TYPE C LECTINS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION NUMBER: US/08/840,062  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: P1019R1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-3881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1487 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
US-08-840-062-7

Query Match 86.0%; Score 43; DB 3; Length 1487;  
Best Local Similarity 77.8%; Pred. No. 31;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
DB 929 CSVPLPSIC 937

RESULT 2

Sun Sep 5 10:36:15 2004

us-09-761-636a-10.open.ra1

```
US-09-489-039A-10842
; Sequence 10842, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10842
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10842

Query Match      82.0%; Score 41; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 SVPLTSVC 9
Db      387 SVPLTSVC 394

RESULT 3
US-09-252-991A-20364
; Sequence 20364, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20364
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20364

Query Match      72.0%; Score 36; DB 4; Length 168;
Best Local Similarity 66.7%; Pred. No. 52;
Matches      6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CSVPLTSVC 9
Db      104 CSVPLTCTC 112

RESULT 4
US-08-688-342-1
; Sequence 1, Application US/08688342
; Patent No. 5871964
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive

Qy      1 CSVPLTSVC 9
Db      104 CSVPLTCTC 112

US-09-489-039A-10842
; Sequence 10842, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10842
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10842

Query Match      82.0%; Score 41; DB 4; Length 422;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 SVPLTSVC 9
Db      387 SVPLTSVC 394

RESULT 3
US-09-252-991A-20364
; Sequence 20364, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20364
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20364

Query Match      72.0%; Score 36; DB 4; Length 168;
Best Local Similarity 66.7%; Pred. No. 52;
Matches      6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CSVPLTSVC 9
Db      104 CSVPLTCTC 112

RESULT 4
US-08-688-342-1
; Sequence 1, Application US/08688342
; Patent No. 5871964
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive

Qy      1 CSVPLTSVC 9
Db      187 CSVPSYSIC 195

Query Match      72.0%; Score 36; DB 2; Length 201;
Best Local Similarity 66.7%; Pred. No. 63;
Matches      6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CSVPLTSVC 9
Db      187 CSVPSYSIC 195

RESULT 5
US-09-113-788-1
; Sequence 1, Application US/09113788
; Patent No. 5969104
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN C-TYPE LECTIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/113,788
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/688,342
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0095-1 CIP
```

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 201 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 IMMEDIATE SOURCE:  
 LIBRARY: MWLRIDT01  
 CLONE: 515847  
 US-09-113-788-1

Query Match 72.0%; Score 36; DB 2; Length 201;  
 Best Local Similarity 66.7%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPLTSVC 9  
 Db 187 CSVPSYSIC 195

## RESULT 6

US-08-586-165-4  
 Sequence 4, Application US/08586165  
 Patent No. 6054298

## GENERAL INFORMATION:

APPLICANT: Laufer, Edward M.  
 APPLICANT: Orozco, Olivia E.  
 APPLICANT: Tabin, Clifford J.

TITLE OF INVENTION: Fringe Proteins and Pattern Formation  
 NUMBER OF SEQUENCES: 9

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 STREET: Two Militia Drive  
 CITY: Lexington  
 STATE: Massachusetts  
 COUNTRY: US  
 ZIP: 02173

## COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/586,165  
 FILING DATE: 16-JAN-1996  
 CLASSIFICATION: 800

## ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia  
 REGISTRATION NUMBER: 32,227  
 REFERENCE/DOCKET NUMBER: HU95-05  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 861-6240  
 TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 4:

## SEQUENCE CHARACTERISTICS:

LENGTH: 359 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FEATURE:

NAME/KEY: CDS

LOCATION: 1..1118

US-08-586-165-4

Query Match 70.0%; Score 35; DB 3; Length 359;  
 Best Local Similarity 44.4%; Pred. No. 1.6e+02;  
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CSVPLTSVC 9  
 Db 293 CSPIWIKIC 301

## RESULT 7

US-09-621-976-4725  
 Sequence 4725, Application US/09621976  
 Patent No. 6639063

## GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.  
 APPLICANT: Jobert, S.  
 APPLICANT: Giordano, J.Y.

TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 FILE REFERENCE: GENSET.054PR2

CURRENT APPLICATION NUMBER: US/09/621,976

CURRENT FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 19335

SOFTWARE: Patent.pm

SEQ ID NO 4725

LENGTH: 134

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: UNSURE

LOCATION: 52

OTHER INFORMATION: Xaa = \*,Tyr

NAME/KEY: UNSURE

LOCATION: 49,51

OTHER INFORMATION: Xaa = Ala,Gly

NAME/KEY: UNSURE

LOCATION: 44

OTHER INFORMATION: Xaa = Cys,Trip

NAME/KEY: UNSURE

LOCATION: 47

OTHER INFORMATION: Xaa = Gly,Arg

NAME/KEY: UNSURE

LOCATION: 53

OTHER INFORMATION: Xaa = Lys,Arg

NAME/KEY: UNSURE

LOCATION: 45

OTHER INFORMATION: Xaa = Phe,Leu

US-09-621-976-4725

Query Match 68.0%; Score 34; DB 4; Length 134;  
 Best Local Similarity 55.6%; Pred. No. 90;  
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CSVPLTSVC 9  
 Db 21 CDVPLYDIC 29

## RESULT 8

US-08-134-012-3

Sequence 3, Application US/08134012

Patent No. 5516652

## GENERAL INFORMATION:

APPLICANT: Abramovitz, Mark

APPLICANT: Boie, Yves

APPLICANT: Grygorczyk, Richard

APPLICANT: Metters, Kathleen

APPLICANT: Rushmore, Thomas H.

APPLICANT: Slipetz, Deborah M.

TITLE OF INVENTION: DNA ENCODING PROSTAGLANDIN RECEPTOR IP

NUMBER OF SEQUENCES: 6

## CORRESPONDENCE ADDRESS:

ADDRESSEE: John Wallen

STREET: 126 E. Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: USA

ZIP: 07065

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/134,012  
FILING DATE: 06-OCT-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 19098  
TELEPHONE: (908) 594-3905  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-134-012-3

Query Match 68.0%; Score 34; DB 1; Length 386;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
Db 251 CSLPLTIRC 259

RESULT 10  
US-09-039-798-3  
Sequence 3, Application US/09039798  
Patent No. 636360  
GENERAL INFORMATION:  
APPLICANT: Abramovitz, Mark  
APPLICANT: Boie, Yves  
APPLICANT: Grygorczyk, Richard  
APPLICANT: Metters, Kathleen  
APPLICANT: Rushmore, Thomas H.  
APPLICANT: Slipetz, Deborah M.  
TITLE OF INVENTION: DNA ENCODING PROSTAGLANDIN RECEPTOR IP  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: J. Mark Hand  
STREET: 126 E. Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/039,798  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/520,519  
FILING DATE: 29-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Hand, J. Mark  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 19098DB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (732) 594-3905  
TELEFAX: (732) 594-4720  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-039-798-3

Query Match 68.0%; Score 34; DB 4; Length 386;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
Db 251 CSLPLTIRC 259

RESULT 11  
US-09-673-395A-479  
Sequence 479, Application US/09673395A

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/134,012  
FILING DATE: 06-OCT-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Wallen, John W  
REGISTRATION NUMBER: 35,403  
REFERENCE/DOCKET NUMBER: 19098  
TELEPHONE: (908) 594-3905  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-134-012-3

Query Match 68.0%; Score 34; DB 1; Length 386;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CSVPLTSVC 9  
Db 251 CSLPLTIRC 259

RESULT 9  
US-08-520-519-3  
Sequence 3, Application US/08520519  
Patent No. 5728808  
GENERAL INFORMATION:  
APPLICANT: Abramovitz, Mark  
APPLICANT: Boie, Yves  
APPLICANT: Grygorczyk, Richard  
APPLICANT: Metters, Kathleen  
APPLICANT: Rushmore, Thomas H.  
APPLICANT: Slipetz, Deborah M.  
TITLE OF INVENTION: DNA ENCODING PROSTAGLANDIN RECEPTOR IP  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jack L. Tribble  
STREET: 126 E. Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/520,519  
FILING DATE: 29-AUG-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Tribble, Jack L.  
REGISTRATION NUMBER: 32,633  
REFERENCE/DOCKET NUMBER: 19098DA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-5321  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

```
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 479
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-479

Query Match      68.0%; Score 34; DB 4; Length 400;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      3 VPLTVC 9
DB      51 LPLTVC 57

RESULT 12
US-08-816-693A-51
; Sequence 51, Application US/08816693A
; Patent No. 5874241
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S
; APPLICANT: Turek, Fred W
; APPLICANT: Pinto, Lawrence H
; TITLE OF INVENTION: Clock Gene and Gene Product
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dressler, Rockey, Milnamow & Katz
; STREET: Two Prudential Plaza, Suite 4700
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,693A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5874241thrup, Thomas E
; REGISTRATION NUMBER: 33,268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5400
; TELEFAX: 312-616-5460
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-816-693A-51

Query Match      68.0%; Score 34; DB 2; Length 747;
Best Local Similarity 66.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 CSVPLTVC 9
DB      215 CRVPLGKVC 223

RESULT 13
US-08-885-291-51
; Sequence 51, Application US/0885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; CURRENT FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816,693
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-885-291-51

Query Match      68.0%; Score 34; DB 3; Length 747;
Best Local Similarity 66.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 CSVPLTVC 9
DB      215 CRVPLGKVC 223

RESULT 14
US-09-496-672-51
; Sequence 51, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-496-672-51

Query Match      68.0%; Score 34; DB 3; Length 747;
Best Local Similarity 66.7%; Pred. No. 5.1e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 CSVPLTVC 9
DB      215 CRVPLGKVC 223

RESULT 15
US-09-402-016A-4
; Sequence 4, Application US/09402016A
```

Patent No. 6610303  
; GENERAL INFORMATION:  
; APPLICANT: HAUSEN, ETHEL-MICHELE DE VILLIERS-ZUR  
; APPLICANT: HAUSEN, HARALD ZUR  
; APPLICANT: LAVERGNE, DONNA  
; APPLICANT: BENTON, CLAIRE  
; TITLE OF INVENTION: PAPILLOMA VIRUS MAIN CAPSID PROTEIN AND THE USE THEREOF IN DIAGNOSIS  
; FILE REFERENCE: 8484-100-999  
; CURRENT APPLICATION NUMBER: US/09/402.016A  
; CURRENT FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: PCT/DE98/00876  
; PRIOR FILING DATE: 1998-03-24  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 126  
; TYPE: PRT  
; ORGANISM: Papilloma virus capsid  
US-09-402-016A-4  
  
Query Match 66.0%; Score 33; DB 4; Length 126;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 CSVPLTS 7  
| | | | |  
Db 54 CKVPLTS 60

Search completed: September 5, 2004, 10:21:59  
Job time : 9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:47:29 ; Search time 5.37374 Seconds  
(without alignments)  
125.302 Million cell updates/sec

Title: US-09-761-636A-11

Perfect score: 42

Sequence: 1 CVPLTSC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Pir1.\*
- 2: Pir2.\*
- 3: Pir3.\*
- 4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	2543	2 T31687	surface antigen - p
2	38	90.5	380	2 D44490	retrovirus-related
3	36	85.7	980	2 T00045	celldextrin phosph
4	36	85.7	2533	2 T28675	alpha-51D immobili
5	36	85.7	2533	2 T28674	alpha-51D-immobili
6	34	81.0	1506	2 T30886	integumentary muc
7	34	81.0	1513	2 A54895	mucin 2, intestina
8	34	81.0	2476	2 T34022	zonadhesin - pig
9	33	78.6	55	2 S66336	protein kinase AK2
10	33	78.6	191	2 I46412	keratin KAP5.4 - s
11	33	78.6	207	2 A64555	hypothetical prote
12	33	78.6	207	2 B71943	hypothetical prote
13	33	78.6	223	2 B38346	ultra-high-sulfur
14	33	78.6	230	2 A38346	ultra-high-sulfur
15	33	78.6	303	1 WZBEM6	gene 20 protein -
16	33	78.6	303	2 C33374	hypothetical prote
17	33	78.6	515	2 T23089	hypothetical prote
18	33	78.6	647	2 T23814	hypothetical prote
19	33	78.6	670	2 F88297	protein M28.1 lim
20	33	78.6	891	2 H86306	F2023.20 protein
21	33	78.6	1227	2 T49563	hypothetical prote
22	33	78.6	3020	2 A43932	hypothetical prote
23	33	78.6	3133	2 S52093	mucin 2 precursor,
24	32	76.2	77	2 JCA4790	hemocytin - silkw
25	32	76.2	270	2 T47421	proteinase inhibit
26	32	76.2	303	2 T42933	hypothetical prote
27	32	76.2	305	2 B53782	hypothetical prote
28	32	76.2	452	2 F84712	peroxisome assembl
29	32	76.2	491	2 A49179	hypothetical prote
					melanoma antigen h

30	32	76.2	662	2 I38400	melanoma-associate
31	32	76.2	668	2 A41234	melanocyte-specifi
32	32	76.2	1319	2 S55598	tegument protein 0
33	32	76.2	2044	2 T13704	still life protein
34	32	76.2	2180	2 T29764	hypothetical prote
35	32	76.2	2395	1 S50820	surface protein ty
36	32	76.2	5376	2 T42215	zonadhesin - mouse
37	31	73.8	63	2 S08572	chymotrypsin/elast
38	31	73.8	110	2 S16496	hypothetical prote
39	31	73.8	120	2 A33787	vascular endotheli
40	31	73.8	128	2 I51295	vascular endotheli
41	31	73.8	146	2 S57956	ovine vascular end
42	31	73.8	161	2 T27849	hypothetical prote
43	31	73.8	169	1 S18946	ultra-high-sulfur
44	31	73.8	182	2 A36686	ultra-high-sulfur
45	31	73.8	190	2 S52130	vascular endotheli

ALIGNMENTS

RESULT 1

T31687  
surface antigen - Paramacium primaurelia  
C:Species: Paramacium primaurelia  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
C:Accession: T31687  
R:Bourgain-Guglielmetti, F.; Caron, F.  
Journal of Eukaryot. Microbiol. 43, 303-314, 1996  
A:Title: Molecular characterization of the D surface protein gene subfamily in Paramacium primaurelia  
A:Reference number: Z21061; MUID:96313351; PMID:8768434  
A:Accession: T31687  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2543 <BOU>  
A:Cross-references: EMBL:X96616; NID:gl235576; PIDN:CAA65436.1  
C:Genetics:  
A:Genetic code: SGC5  
C:Superfamily: G surface protein

Query Match 92.9%; Score 39; DB 2; Length 2543;  
Best Local Similarity 71.4%; Pred.No. 26;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
|:|:|  
Db 2158 CIPITSC 2164

RESULT 2

D44490  
retrovirus-related reverse transcriptase homolog (clone NVC) - pteromalid wasp (Nasonia vitripennis)  
C:Species: Nasonia vitripennis  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 01-Dec-2000  
C:Accession: D44490; F40442  
R:Burke, W.D.; Eickbush, D.G.; Xiong, Y.; Jakubczak, J.; Eickbush, T.H.  
Mol. Biol. Evol. 10, 163-185, 1993  
A:Title: Sequence relationship of retrotransposable elements R1 and R2 within and between species of the genus Nasonia  
A:Reference number: A44490; MUID:93196484; PMID:8383793  
A:Contents: retrotransposable element R1  
A:Accession: D44490  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-380 <BUR>  
R:Jakubczak, J.L.; Burke, W.D.; Eickbush, T.H.  
Proc. Natl. Acad. Sci. U.S.A. 88, 3295-3299, 1991  
A:Title: Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects.  
A:Reference number: A40442; MUID:91195337; PMID:1849649  
A:Accession: F40442  
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 304-320 <JAK>

C:Superfamily: silkworm pol protein

Query Match 90.5%; Score 38; DB 2; Length 380;  
Best Local Similarity 85.7%; Pred. No. 8.1;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 356 CVPLSSC 362

RESULT 3

T00045  
cellocodextrin phosphorylase - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 15-Oct-1999  
C:Accession: T00045  
R:Kawaguchi, T.; Ikeuchi, Y.; Tsutsumi, N.; Kan, A.; Sumitani, J.; Arai, M.  
J. Ferment. Bioeng. 85, 144-149, 1998  
A:Title: Cloning, nucleotide sequence, and expression of the Clostridium thermocellum cel  
A:Reference number: Z14077  
A:Accession: T00045  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-980 <KAW>  
A:Cross-references: EMBL:AB006822; NID:dl117395; PIDN:BA022081.1; PID:dl022940  
A:Experimental source: ATCC 27405  
C:Genetics:  
A:Gene: cdp

Query Match 85.7%; Score 36; DB 2; Length 980;  
Best Local Similarity 57.1%; Pred. No. 44;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 550 CIPMTAC 556

RESULT 4

T28675  
alpha-51D immobilization antigen - Paramesidium tetraurelia  
C:Species: Paramesidium tetraurelia  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T28675  
R:Schwegmann, K.J.  
Submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20506  
A:Accession: T28675  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2533 <SCH>  
A:Cross-references: EMBL:X96400; PIDN:CAA65264.1  
C:Genetics:  
A:Gene: alpha-51D  
A:Genetic code: SGC5  
A:Introns: 280/3; 538/2; 1248/2  
C:Superfamily: G surface protein

Query Match 85.7%; Score 36; DB 2; Length 2533;  
Best Local Similarity 57.1%; Pred. No. 98;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 2149 CIPITNC 2155

RESULT 5

T28674  
alpha-51D-immobilization antigen - Paramesidium tetraurelia  
C:Species: Paramesidium tetraurelia  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000

C:Accession: T28674  
R:Schmidt, H.J.  
Submitted to the EMBL Data Library, March 1995  
A:Reference number: Z20505  
A:Accession: T28674  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2533 <SCH>  
A:Cross-references: EMBL:X85135; NID:g728634; PID:g728635; PIDN:CAA59447.1  
C:Genetics:  
A:Genetic code: SGC5  
A:Note: alpha-51D  
C:Superfamily: G surface protein

Query Match 85.7%; Score 36; DB 2; Length 2533;  
Best Local Similarity 57.1%; Pred. No. 98;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 2149 CIPITNC 2155

RESULT 6

T30886  
integumentary mucin B.1 - African clawed frog (fragment)  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 03-Nov-2000  
C:Accession: T30886  
R:Joba, W.; Hoffmann, W.  
J. Biol. Chem. 272, 1805-1810, 1997  
A:Title: Similarities of integumentary mucin B.1 (FIM-B.1) from Xenopus laevis and prep  
A:Reference number: Z20920; MUID:97153143; PMID:8999864  
A:Accession: T30886  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1506 <JOB>  
A:Cross-references: EMBL:Y08296; NID:g1839051; PIDN:CAA69604.1; PID:gl839052  
C:Genetics:  
A:Note: FIM-B.1  
C:Superfamily: pig submaxillary mucin

Query Match 81.0%; Score 34; DB 2; Length 1506;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 1292 CVPLSKC 1298

RESULT 7

A54895  
mucin 2, intestinal/tracheal - rat (fragment)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 11-Jan-2000  
C:Accession: A54895  
R:Ohmori, H.; Dohrman, A.F.; Gallup, M.; Tsuda, T.; Kai, H.; Gum Jr., J.R.; Kim, Y.S.;  
J. Biol. Chem. 269, 17833-17840, 1994  
A:Title: Molecular cloning of the amino-terminal region of a rat MUC 2 mucin gene homo  
A:Reference number: A54895; MUID:94299489; PMID:8027037  
A:Accession: A54895  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1513 <OHM>  
A:Cross-references: GB:U07615  
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; vo  
C:Keywords: intestine

Query Match 81.0%; Score 34; DB 2; Length 1513;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



QY 1 CVPLTSC 7  
Db 709 CVPLSKC 715

RESULT 8  
T34022  
zonadhesin - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T34022  
R:Hardy, D.M.; Garbers, D.L.  
J. Biol. Chem. 270, 26025-26028, 1995  
A:Title: A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix  
A:Reference number: Z21464; MUID:96064658; PMID:7592795  
A:Accession: T34022  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-2476 <HAR>  
A:Cross-references: EMBL:U40024; NID:gl066465; PID:gl066466; PIDN:AAC48486.1  
A:Experimental source: strain Weishan; testis  
C:Genetics:  
A:Gene: Zan  
C:Function:  
A:Description: may be involved in sperm adhesion to the zona pellucida

Query Match 81.0%; Score 34; DB 2; Length 2476;  
Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
Db 1505 CVPLSQC 1511

RESULT 9  
S66336  
protein kinase AK23 (EC 2.7.1.1) - Arabidopsis thaliana (fragment)  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 24-Sep-1999  
C:Accession: S66336; S58262  
R:Thuenmler, F.; Kirchner, M.; Teuber, R.; Dittrich, P.  
Plant Mol. Biol. 29, 551-565, 1995  
A:Title: Differential accumulation of the transcripts of 22 novel protein kinase genes in Arabidopsis thaliana  
A:Reference number: S66314; MUID:96123233; PMID:8534852  
A:Accession: S66336  
A:Molecule type: DNA  
A:Residues: 1-55 <THU>  
A:Cross-references: EMBL:X86968; NID:g928913; PIDN:CAA60531.1; PID:g928914  
C:Genetics:  
A:Gene: AK23  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homologous  
C:Keywords: ATP; phosphotransferase; protein kinase  
F:1-55/Domain: protein kinase homology (fragment) <KIN>

Query Match 78.6%; Score 33; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTS 6  
Db 27 CVPLTS 32

RESULT 10  
I46412  
keratin KAP5.4 - sheep  
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)  
C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 24-Sep-1999  
C:Accession: I46412; S34215  
R:Jenkins, B.J.; Powell, B.C.  
J. Invest. Dermatol. 103, 310-317, 1994  
A:Title: Differential expression of genes encoding a cysteine-rich keratin family in the

A:Reference number: I46412; MUID:94358466; PMID:7521375  
A:Accession: I46412  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-191 <JEN>  
A:Cross-references: EMBL:X73434; NID:g313719; PIDN:CAA51829.1; PID:g313720  
C:Genetics:  
A:Gene: KRTAP5.4  
C:Superfamily: ultra-high-sulfur keratin

Query Match 78.6%; Score 33; DB 2; Length 191;  
Best Local Similarity 71.4%; Pred. No. 40;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
Db 91 CVPVCSC 97

RESULT 11  
A64655  
hypothetical protein HP1081 - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: A64655  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: A64655  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-207 <TOM>  
A:Cross-references: GB:AE000615; GB:AE000511; NID:g2314230; PIDN:AAD08133.1; PID:g23142

Query Match 78.6%; Score 33; DB 2; Length 207;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
Db 134 CVPQTTC 140

RESULT 12  
B71943  
hypothetical protein jhp0344 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: B71943  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: B71943  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-207 <ARN>  
A:Cross-references: GB:AE001470; GB:AE001439; NID:g4154869; PIDN:AAD05923.1; PID:g415487  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0344

Query Match 78.6%; Score 33; DB 2; Length 207;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7

Sun Sep 5 10:36:19 2004

us-09-761-636a-11.open.rpr

Db 134 CVPQTAC 140  
||| |:  
||| |:

## RESULT 13

B38346  
A;Title: ultra-high-sulfur keratin 2 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
C;Accession: A38660; B38346  
R;Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.  
J. Biol. Chem. 266, 4024, 1991  
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin  
A;Reference number: A38660; MUID:91154184; PMID:1840598  
A;Accession: A38660  
A;Molecule type: DNA  
A;Residues: 1-223 <WO2>  
A;Cross-references: GB:M37760; NID:g200963; PIDN:AAA40107.1; PID:g200964  
A;Note: this is a correction  
R;Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.  
J. Biol. Chem. 265, 21375-21380, 1990  
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin  
A;Reference number: A38346; MUID:91065960; PMID:2250030  
A;Accession: B38346  
A;Molecule type: DNA  
A;Residues: 1-21, 'GGCGGGCGGCGGCGSSCKPVCC', 22-40, 'GSS', 44-45, 'G', 47-48, 'S', 50, 'GSS',  
<WOO>  
A;Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962  
A;Note: the sequence reported in this paper has been corrected. See A38660  
C;Superfamily: ultra-high-sulfur keratin

Query Match 78.6%; Score 33; DB 2; Length 223;  
Best Local Similarity 71.4%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
|||:  
|||:  
Db 28 CVPVCSC 34

## RESULT 14

A38346  
A;Title: ultra-high-sulfur keratin 1 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 24-Sep-1999  
C;Accession: A38346  
R;Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.  
J. Biol. Chem. 265, 21375-21380, 1990  
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin  
A;Reference number: A38346; MUID:91065960; PMID:2250030  
A;Accession: A38346  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-230 <WOO>  
A;Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962  
C;Superfamily: ultra-high-sulfur keratin

Query Match 78.6%; Score 33; DB 2; Length 230;  
Best Local Similarity 71.4%; Pred. No. 47;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
|||:  
|||:  
Db 55 CVPVCSC 61

## RESULT 15

WZBEM6  
gene 20 protein - saimirine herpesvirus 1 (strain 11)  
C;Species: saimirine herpesvirus 1  
A;Note: host Saimiri sciureus (common squirrel monkey)  
C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 16-Jul-1999  
C;Accession: B36808

R;Albrecht, J.  
submitted to the EMBL Data Library, January 1992  
A;Description: Primary structure of the herpesvirus saimiri genome.  
A;Reference number: A36806  
A;Accession: B36808  
A;Molecule type: DNA  
A;Residues: 1-303 <ALB>  
A;Cross-references: GB:X64346; NID:g60320; PIDN:CAA45644.1; PID:g60342  
R;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W  
J. Virol. 66, 5047-5058, 1992  
A;Title: Primary structure of the herpesvirus saimiri genome.  
A;Reference number: A37309; MUID:92333688; PMID:1321287  
A;Contents: annotation; possible protein-coding frames  
A;Note: neither amino acid nor nucleotide sequence is given  
C;Genetics:  
A;Gene: 20  
C;Superfamily: varicella-zoster virus gene 35 protein

Query Match 78.6%; Score 33; DB 1; Length 303;  
Best Local Similarity 85.7%; Pred. No. 60;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
|||:  
|||:  
Db 75 CVMULTSC 81

Search completed: September 5, 2004, 10:01:21  
Job time : 6.37374 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:38:39 ; Search time 3.18182 Seconds  
(without alignments)  
114.554 Million cell updates/sec

Title: US-09-761-636A-11  
Perfect score: 42  
Sequence: 1 CVPLTSC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	38	90.5	383	1 POL3 NASVI
2	35	83.3	727	1 PKPI_BOVIN
3	34	81.0	1513	1 MUC2_RAT
4	34	81.0	2476	1 ZAN_FIG
5	33	78.6	251	1 K10A_HUMAN
6	33	78.6	303	1 UL24_HSVSA
7	33	78.6	471	1 FBX3_HUMAN
8	33	78.6	515	1 KE4L_CAEEL
9	33	78.6	2282	1 ZAN_RABIT
10	33	78.6	3133	1 HMCT_BOMMO
11	33	78.6	5179	1 MUC2_HUMAN
12	32	76.2	77	1 FPIF_BOMMO
13	32	76.2	305	1 PEX2_RAT
14	32	76.2	379	1 PSPB_DICDI
15	32	76.2	491	1 PM17_BOVIN
16	32	76.2	661	1 PM17_HUMAN
17	32	76.2	2044	1 SIF2_DROME
18	32	76.2	5376	1 ZAN_MOUSE
19	31	73.8	65	1 ICE2_ASCSU
20	31	73.8	146	1 VEGA_SHEEP
21	31	73.8	164	1 VEGA_CAVPO
22	31	73.8	169	1 KRUA_HUMAN
23	31	73.8	182	1 KRUC_SHEEP
24	31	73.8	190	1 VEGA_BOVIN
25	31	73.8	190	1 VEGA_HORSE
26	31	73.8	190	1 VEGA_MESAU
27	31	73.8	190	1 VEGA_FIG
28	31	73.8	194	1 KRUB_HUMAN
29	31	73.8	200	1 RISA_CHLPN
30	31	73.8	214	1 VEGA_CANFA
31	31	73.8	214	1 VEGA_MOUSE
32	31	73.8	214	1 VEGA_RAT
33	31	73.8	216	1 VEGA_CHICK

34	31	73.8	232	1 VEGA_HUMAN	P15692 homo sapien
35	31	73.8	256	1 YQSC_ECOLI	Q46809 escherichia
36	31	73.8	289	1 TNR5_MOUSE	P27512 mus musculus
37	31	73.8	346	1 GATD_ECOLI	P37190 escherichia
38	31	73.8	468	1 T10A_HUMAN	O00220 homo sapien
39	31	73.8	1576	1 YLK3_CAEEL	P41951 caenorhabdi
40	31	73.8	2704	1 G168_FARPR	P17053 paramacium
41	31	73.8	2715	1 G156_FARPR	P13837 paramacium
42	31	73.8	2812	1 ZAN_HUMAN	Q9V493 homo sapien
43	30	71.4	57	1 6H9A_HUMAN	Q13653 homo sapien
44	30	71.4	157	1 DIF_DICDI	P08797 dictyosteli
45	30	71.4	166	1 LITE_HUMAN	P48304 homo sapien

ALIGNMENTS

RESULT 1  
POL3 NASVI  
ID POL3 NASVI STANDARD; PRT; 383 AA.  
AC Q03271;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Retrovirus-related POL polyprotein from type I retrotransposable  
DE element R1 [Contains: Reverse transcriptase (EC 2.7.7.49);  
DE Endonuclease] (Fragment).  
OS Nasonia vitripennis (Parasitic wasp).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Chalcidoidea;  
OC Pteromalidae; Nasonia.  
OX NCBI\_TaxID=7425;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93196484; PubMed=8383793;  
RA Burke W.D., Eickbush D.G., Xiong Y., Jakubczak J.L., Eickbush T.H.;  
RT "Sequence relationship of retrotransposable elements R1 and R2 within  
RT and between divergent insect species";  
RL Mol. Biol. Evol. 10:163-185(1993).  
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA} (N).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L00942; AAA30339.1; -.  
DR PIR; D44490; D44490.  
DR InterPro; IPR000477; RVTse.  
DR Pfam; PF00078; rvt; 1.  
KW Transferase; RNA-directed DNA polymerase; Transposable element;  
KW Hydrolase; Nuclease; Endonuclease.  
FT NON\_TER 1  
FT DOMAIN <1 228 REVERSE TRANSCRIPTASE  
FT DOMAIN 229 383 NUCLEIC ACID-BINDING ENDONUCLEASE.  
SQ SEQUENCE 383 AA; 43411 MW; DE296B380925251B CRC64;  
Query Match 90.5%; Score 38; DB 1; Length 383;  
Best Local Similarity 85.7%; Pred. No. 2.2;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
Db 359 CVPLSSC 365  
RESULT 2  
PKPI\_BOVIN  
ID PKPI\_BOVIN STANDARD; PRT; 727 AA.

Q28161;  
 AC 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Plakophilin 1 (Band-6-protein) (B6P).  
 GN PKP1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=95196971; PubMed=7890138;  
 RX Heid H.W., Schmidt A., Zimbelmann R., Schaefer S.,  
 RA Winter-Simanowski S., Stump S., Keith M., Figge U., Schnolzer M.,  
 RA Franke W.W.;  
 RA "Cell type-specific desmosomal plaque proteins of the plakoglobin  
 RT family: plakophilin 1 (band 6 protein).";  
 RT Differentiation 58:113-131(1994).  
 RL  
 CC -!- FUNCTION: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY  
 CC -!- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes (By  
 CC -!- Similarity).  
 CC -!- SIMILARITY: Belongs to the beta-catenin family.  
 CC -!- SIMILARITY: Contains 8 ARM repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; Z37975; CAA86029.1; -  
 DR GO; GO:0030057; C:desmosome; ISS.  
 DR GO; GO:0005634; C:nucleus; ISS.  
 DR GO; GO:0019215; F:intermediate filament binding; ISS.  
 DR GO; GO:0030280; F:structural constituent of epidermis; ISS.  
 DR GO; GO:0007155; P:cell adhesion; ISS.  
 DR GO; GO:0007165; P:signal transduction; ISS.  
 DR InterPro; IPR008938; ARM.  
 DR InterPro; IPR000225; Armadillo.  
 DR Pfam; PF00514; Armadillo\_seg; 4.  
 DR SMART; SM00185; ARM; 5.  
 DR PROSITE; PS0176; ARM\_REPEAT; 3.  
 DR Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;  
 KW Repeat.  
 FT REPEAT 235 275 ARM 1.  
 FT REPEAT 278 317 ARM 2.  
 FT REPEAT 320 360 ARM 3.  
 FT REPEAT 419 464 ARM 4.  
 FT REPEAT 517 557 ARM 5.  
 FT REPEAT 565 604 ARM 6.  
 FT REPEAT 606 650 ARM 7.  
 FT REPEAT 653 693 ARM 8.  
 SQ SEQUENCE 727 AA; 80180 MW; 3A27979279BCCBDF CRC64;  
 Query Match 83.3%; Score 35; DB 1; Length 727;  
 Best Local Similarity 85.7%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVELTSC 7  
 Db 215 CVELTSC 221  
 RESULT 3  
 MUC2\_RAT  
 ID MUC2\_RAT STANDARD; PRT; 1513 AA.  
 AC Q2635;  
 DT 28-FEB-2003 (Rel. 41, Created)

28-FEB-2003 (Rel. 41, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Mucin 2 precursor (Intestinal mucin 2) (Fragment).  
 MUC2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUR=Intestine;  
 RX MEDLINE=94299489; PubMed=8027037;  
 RA Ohmori H., Dohrman A.F., Gallup M., Tsuda T., Kai H., Gum J.R. Jr.,  
 RA Kim Y.S., Basbaum C.B.;  
 RA "Molecular cloning of the amino-terminal region of a rat MUC 2 mucin  
 RT gene homologue. Evidence for expression in both intestine and  
 RT airway.";  
 RT J. Biol. Chem. 269:17833-17840(1994).  
 RN [2]  
 RN SEQUENCE OF 21-36.  
 RP TISSUR=Intestinal epithelium;  
 RX MEDLINE=98180965; PubMed=9512496;  
 RA Khatri I.A., Forstner G.G., Forstner J.F.;  
 RA "Susceptibility of the cysteine-rich N-terminal and C-terminal ends of  
 RT rat intestinal mucin Muc 2 to proteolytic cleavage.";  
 RL Biochem. J. 331:323-330(1998).  
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and  
 CC other mucus membrane-containing organs. Thought to provide a  
 CC protective, lubricating barrier against particles and infectious  
 CC agents at mucosal surfaces.  
 CC -!- SUBUNIT: Multimeric.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Expressed in intestine and airway.  
 CC -!- SIMILARITY: Contains 1 VWFC domain.  
 CC -!- SIMILARITY: Contains 3 VWFC domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; U07615; AAA21655.2; -  
 DR PIR; A54895; A54895.  
 DR InterPro; IPR002919; TIL\_Cysrich.  
 DR InterPro; IPR001007; VWFC.  
 DR InterPro; IPR001846; VWFC\_D.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00094; vwd; 3.  
 DR SMART; SM00214; VMC; 1.  
 DR SMART; SM00216; VMD; 3.  
 DR PROSITE; PS01208; VWFC\_1; FALSE\_NEG.  
 DR PROSITE; PS0184; VWFC\_2; FALSE\_NEG.  
 KW Signal; Repeat; Glycoprotein.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 >1513 MUCIN 2.  
 FT DOMAIN 21 182 VWFC 1.  
 FT DOMAIN 350 410 VWFC.  
 FT DOMAIN 377 541 VWFC 2.  
 FT DOMAIN 847 1005 VWFC 3.  
 FT DOMAIN 1392 >1513 APPROXIMATE REPEATS.  
 FT REPEAT 1392 1407 1.  
 FT REPEAT 1408 1423 2.  
 FT REPEAT 1424 1434 3.  
 FT REPEAT 1435 1445 4.  
 FT REPEAT 1446 1456 5.  
 FT REPEAT 1457 1467 6.  
 FT REPEAT 1468 1478 7.  
 FT REPEAT 1479 1489 8.  
 FT REPEAT 1490 1500 9.  
 FT REPEAT 1501 1511 10.

```

FT REPEAT 1512 >1513 11.
FT CARBOHYD 160 160 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 767 767 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 837 837 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 892 892 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 1151 1151 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 1227 1227 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 1350 1350 N-LINKED (GLCNAC. .) (POTENTIAL)
FT NON TER 1513 1513
SQ SEQUENCE 1513 AA; 26109DCA1BE7D008 CRC64;

Query Match 81.0%; Score 34; DB 1; Length 1513;
Best Local Similarity 71.4%; Pred. No. 56;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 709 CVPLSKC 715

RESULT 4
ZAN_PIG STANDARD; PRT; 2476 AA.
AC Q28983;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 823-830; 859-872; 893-890;
RP 920-925; 960-967; 1235-1244; 1349-1354; 1518-1532; 1624-1656;
RP 1658-1667; 1777-1795 AND 1914-1921.
RC STRAIN=Meishan; TISSUE=Testis;
RX MEDLINE=96064658; PubMed=7592795;
RA Hardy D.M., Garbers D.L.;
RT "A sperm membrane protein that binds in a species-specific manner to
RT the egg extracellular matrix is homologous to von Willebrand
RT factor".
RL J. Biol. Chem. 270:26025-26028(1995).
CC -!- FUNCTION: Binds in a species-specific manner to the zona pellucida
CC of the egg. May be involved in gamete recognition and/or
CC signaling.
CC -!- SUBUNIT: Probably forms covalent oligomers.
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein, exclusively on the
CC apical region of the sperm head (By similarity).
CC -!- TISSUE SPECIFICITY: In testis, primarily in haploid spermatids.
CC Not in lung, liver, heart, spleen, brain, kidney, epididymis.
CC -!- DOMAIN: The MAM domains probably mediate sperm adhesion to the
CC zona pellucida.
CC -!- DOMAIN: During sperm migration through the reproductive tracts,
CC the mucin-like domain might inhibit inappropriate trapping of
CC spermatozoa or promoting adhesion to the oviductal isthmus.
CC -!- DOMAIN: The VWFD domains 2 and 3 may mediate covalent
CC oligomerization (By similarity to human intestinal mucin MUC2).
CC -!- PTM: THE MAM DOMAINS AND THE MUCIN-LIKE DOMAINS ARE MISSING
CC FROM THE ZONADHESIN THAT BINDS TO THE EGG EXTRACELLULAR MATRIX.
CC PROCESSING MIGHT OCCUR DURING SPERM MATURATION AND/OR
CC CAPACITATION.
CC -!- SIMILARITY: Contains 2 MAM domains.
CC -!- SIMILARITY: Contains 5 VWFD domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U40024; AAC48486.1; -.
CC FIR; T34022; T34022.
CC HSSP; P56682; ICCV.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR000998; MAM_domain.
CC InterPro; IPR002919; TIL_Cysrich.
CC InterPro; IPR003328; Tila_Cysrich.
CC InterPro; IPR001007; VWF_C.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00629; MAM; 2.
CC Pfam; PF01826; TIL; 5.
CC Pfam; PF02345; Tila; 5.
CC Pfam; PF00094; vwd; 4.
CC SMART; SM00137; MAM; 1.
CC SMART; SM00214; VWC; 2.
CC SMART; SMO0216; VWD; 4.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 4.
CC PROSITE; PS00026; EGF_3; 1.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS00060; MAM_2; 2.
CC Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
CC Repeat.
CC SIGNAL 1 29
CC CHAIN 30 2476
CC DOMAIN 30 2418
CC TRANSMEM 2419 2439
CC DOMAIN 2440 2476
CC DOMAIN 31 144
CC DOMAIN 147 312
CC DOMAIN 319 687
CC DOMAIN 688 799
CC DOMAIN 800 1184
CC DOMAIN 1185 1573
CC DOMAIN 1574 1968
CC DOMAIN 1969 2370
CC DOMAIN 2366 2402
CC DISULFID 2370 2381
CC DISULFID 2375 2390
CC DISULFID 2392 2401
CC CARBOHYD 109 109
CC CARBOHYD 269 269
CC CARBOHYD 735 735
CC CARBOHYD 758 758
CC CARBOHYD 833 833
CC CARBOHYD 1154 1154
CC CARBOHYD 1329 1329
CC CARBOHYD 1448 1448
CC CARBOHYD 1544 1544
CC CARBOHYD 1596 1596
CC CARBOHYD 1654 1654
CC CARBOHYD 1843 1843
CC CARBOHYD 1965 1965
CC CARBOHYD 2122 2122
CC CARBOHYD 2165 2165
CC CARBOHYD 2178 2178
CC CARBOHYD 2329 2329
CC CARBOHYD 2359 2359
CC CONFLICT 823 823
CC CONFLICT 923 923
CC CONFLICT 965 965
CC CONFLICT 1241 1241
CC SEQUENCE 2476 AA; 270364 MW; A13B690375A6548C CRC64;

```

```

Query Match      81.0%; Score 34; DB 1; Length 2476;
Best Local Similarity 71.4%; Pred. No. 92;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
Db 1505 CVPLSQC 1511

RESULT 5
K10A HUMAN
ID K10A_HUMAN STANDARD; PRT; 251 AA.
AC P60014;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Keratin associated protein KAP10-10 (Keratin associated protein 10.10)
DE High sulfur keratin associated protein 10.10.
DE KRTAP10-10 OR KRTAP10.10 OR KAP10.10.
GN KRTAP10-10 OR KRTAP10.10 OR KAP10.10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Scalp;
RA Rogers M.A., Langbein L., Winter H., Beckmann I., Praetzel S.,
RA Schweizer J.;
RT "Keratin associated proteins (KAPs): Characterisation of a second high
RT sulfur KAP gene domain on human chromosome 21.";
RL J. Invest. Dermatol. 0:0-0(2003).
CC -!- FUNCTION: In the hair cortex, hair keratin intermediate filaments
CC are embedded in an interfilamentous matrix, consisting of hair
CC keratin-associated protein (KRTAP), which are essential for the
CC formation of a rigid and resistant hair shaft through their
CC extensive disulfide bond cross-linking with abundant cysteine
CC residues of hair keratins. The matrix proteins include the high-
CC sulfur and high-glycine-tyrosine keratins.
CC -!- SUBUNIT: Interacts with hair keratins.
CC -!- TISSUE SPECIFICITY: Expressed exclusively in a narrow region of
CC the middle portion of hair fiber cuticle.
CC -!- SIMILARITY: Belongs to the KRTAP type 10 family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

-----
DR EMBL; AJ566387; CAD97467.1; -
DR EMBL; AL773602; -; NOT ANNOTATED_CDS.
DR Genew; HGNC:22972; KRTAP10-10.
KW Keratin; Repeat; Multigene family.
FT DOMAIN 26 224 15 X 5 AA REPEATS OF C-C-[EAQTVR]-
[SPTQA]-[SCAV].
FT REPEAT 26 30 1.
FT REPEAT 31 35 2.
FT REPEAT 32 35 2.
FT REPEAT 52 56 3.
FT REPEAT 84 88 4.
FT REPEAT 94 98 5.
FT REPEAT 99 103 6.
FT REPEAT 104 109 7.
FT REPEAT 126 130 8.
FT REPEAT 136 140 9.
FT REPEAT 146 150 10.
FT REPEAT 168 172 11.
FT REPEAT 178 182 12.
FT REPEAT 183 187 13.
FT REPEAT 202 206 14.
FT REPEAT 220 224 15.
FT SEQUENCE 251 AA; 25570 MW; A128CDBB45B74962 CRC64;

Query Match      78.6%; Score 33; DB 1; Length 251;
Best Local Similarity 71.4%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
Db 203 CVPVSC 209

RESULT 6
UL24 HSVSA
ID UL24_HSVSA STANDARD; PRT; 303 AA.
AC Q01005;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Gene 20 protein.
DE Gene 20.
GN Herpesvirus saimiri (strain 11).
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333686; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL24,
CC EBV-1 37, EBV BXRF1, HCMV UL76, ILTV ORF3, AND VZV 35.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

-----
DR EMBL; X64346; CAA45644.1; -
DR InterPro; IPR002580; Herpes_UL24.
DR Pfam; PF01646; Herpes_UL24; 1.
SQ SEQUENCE 303 AA; 34942 MW; DP6D59F7A1C83A0B CRC64;

Query Match      78.6%; Score 33; DB 1; Length 303;
Best Local Similarity 85.7%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
Db 75 CVMLTSC 81

RESULT 7
FBX3 HUMAN
ID FBX3_HUMAN STANDARD; PRT; 471 AA.
AC Q9UK99; Q9NUX2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE F-box only protein 3.
DE FBX03 OR FBX3.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
```

RA Nishikawa T., Negai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
RA Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.,  
RT "NEDO human cDNA sequencing project.",  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 6-415 FROM N.A.  
RX MEDLINE=20003061; PubMed=10531037;  
RA Winston J.T., Koepf D.M., Zhu C., Ellledge S.J., Harper J.W.;  
RT "A family of mammalian F-box proteins.",  
RL Curr. Biol. 9:1180-1182(1999)  
CC -!- FUNCTION: Probably recognizes and binds to some phosphorylated  
CC proteins and promotes their ubiquitination and degradation.  
CC -!- SUBUNIT: Part of a SCF (SKP1-cullin-F-box) protein ligase complex  
CC (By similarity).  
CC -!- SIMILARITY: Contains 1 F-box domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AK001943; BAA91991.1; -;  
DR EMBL; AF176702; AAF03702.1; -;  
DR Genew; HGNC:13582; FBXO3.  
DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; TAS.  
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; TAS.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
DR InterPro; IPR007474; DUF525.  
DR InterPro; IPR001810; F-box.  
DR InterPro; IPR008945; Skp1\_Skp2.  
DR Pfam; PF04379; DUF525; 1.  
DR SMART; SM00256; FBXO; 1.  
DR PROSITE; PSS0181; FBXO; 1.  
KW Ub1 conjugation pathway.  
FT DOMAIN 10 56 F-BOX.  
FT DOMAIN 419 452 ASP/GLU-RICH (HIGHLY ACIDIC).  
FT DOMAIN 453 456 POLY-ARG.  
FT DOMAIN 463 466 POLY-ARG.  
FT CONFLICT 164 164 T -> A (IN REF. 2).  
FT CONFLICT 414 415 EM -> VS (IN REF. 2).  
SQ SEQUENCE 471 AA; 54590 MW; F7AA88193E14E67E CRC64;  
  
Query Match 78.6%; Score 33; DB 1; Length 471;  
Best Local Similarity 71.4%; Pred. No. 28;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTSC 7  
Db 193 CLPLTFC 199  
  
RESULT 8  
KB4L\_CAEEL  
ID KE4L CAEEL STANDARD; PRT; 515 AA.  
AC Q9XTQ7;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hypothetical Ke4-like protein H13N06.5 in chromosome X.  
GN H13N06.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Lennard N.;

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family.  
CC KE4/Catsup subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; Z99942; CAB17070.1; -;  
DR PIR; T23089; T23089.  
DR WormPep; H13N06.5; CE18815.  
DR InterPro; IPR003689; Zip\_1\_transpt\_Zip.  
DR Pfam; PF02535; Zip; 1.  
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein.  
FT TRANSMEM 27 47 POTENTIAL.  
FT TRANSMEM 49 69 POTENTIAL.  
FT TRANSMEM 214 234 POTENTIAL.  
FT TRANSMEM 247 267 POTENTIAL.  
FT TRANSMEM 297 317 POTENTIAL.  
FT TRANSMEM 386 406 POTENTIAL.  
FT TRANSMEM 429 449 POTENTIAL.  
FT TRANSMEM 463 483 POTENTIAL.  
FT DOMAIN 92 182 HIS-RICH.  
FT CARBOHYD 7 7 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 379 379 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 488 488 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 515 AA; 55500 MW; 17D7E854F4E1DAAF CRC64;  
  
Query Match 78.6%; Score 33; DB 1; Length 515;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTSC 7  
Db 24 CLPLTSC 30  
  
RESULT 9  
ZAN\_RABIT  
ID ZAN RABIT STANDARD; PRT; 2382 AA.  
AC P57999;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zonadhesin (Fragment).  
GN ZAN.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Testis;  
RX MEDLINE=21573741; PubMed=11717130;  
RA Lea I.A., Sivashanmugam P., O'Rand M.G.;  
RT "Zonadhesin: characterization, localization, and zona pellucida  
RT binding.",  
RL Biol. Reprod. 65:1691-1700(2001).  
CC -!- FUNCTION: Binds in a species-specific manner to the zona pellucida  
CC of the egg. May be involved in gamete recognition and/or  
CC signaling (By similarity).  
CC -!- SUBUNIT: Probably forms covalent oligomers (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein, exclusively on the  
CC apical region of the sperm head (By similarity).  
CC -!- DOMAIN: The NAM domains probably mediate sperm adhesion to the  
CC zona pellucida.  
CC -!- DOMAIN: During sperm migration through the reproductive tracts,

Best Local Similarity 71.4%; Pred. No. 1.4e+02; Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CYP1TSC 7  
Db 1316 CYP1TSC 1322

RESULT 10  
HMCT\_BOMMO STANDARD; PRT; 3133 AA.  
ID HMCT\_BOMMO  
AC P38092;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Hemocytin precursor (Humoral lectin).  
OS Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;  
OC Bombycidae; Bombyx.  
OX NCBI\_TaxID=7091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fuyou x Tokai; TISSUE=Hemocyte;  
RX MEDLINE=95178544; PubMed=7873598;  
RA Kotani E., Yamakawa M., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,  
RA Matsubara F., Tanai K., Kadono-Okuda K., Kato Y., Mori H.;  
RA "Cloning and expression of the gene of hemocytin, an insect humoral  
RT lectin which is homologous with the mammalian von Willebrand  
RT factor.";  
RL Biochim. Biophys. Acta 1260:245-258 (1995).  
RN [2]  
RP SEQUENCE OF 2221-3133 FROM N.A.  
RA Kotani E., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,  
RA Matsubara F., Yamakawa M.;  
RL submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Adhesive protein and relates to hemostasis or  
CC encapsulation of foreign substances for self-defense.  
CC -!- DEVELOPMENTAL STAGE: Expressed in hemocytes during larval-pupal  
CC metamorphosis.  
CC -!- INDUCTION: Hemagglutination activity is increased by bacterial  
CC or viral infection and inhibited by D-mannose, N-acetyl-D-  
CC galactosamine and D-maltose.  
CC -!- PTM: May be converted into the 260 kDa mature hemocytin by  
CC proteolysis.  
CC -!- SIMILARITY: TO MAMMALIAN VON WILLEBRAND FACTOR.  
CC -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT  
CC OF HUMAN MUCIN 2.  
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.  
CC -!- SIMILARITY: Contains 2 WFCC domains.  
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; D29738; BAA06160.1; -;  
CC EMBL; D14035; BAA03124.1; -;  
CC PIR; S52093; S52093.  
CC HSSP; P12259; 1C2T.  
CC InterPro; IPR006207; Cys\_knot\_C.  
CC InterPro; IPR000421; FAS8\_C.  
CC InterPro; IPR008979; Gal\_bind\_like.  
CC InterPro; IPR002172; LDL\_receptor\_A.  
CC InterPro; IPR002919; TIL\_Cysrich.  
CC InterPro; IPR001007; VWF\_C.  
CC InterPro; IPR001846; VWF\_D.  
CC Pfam; PF00754; F5\_F8\_type\_C; 2.  
CC Pfam; PF01826; TIL; 6.  
CC

the mucin-like domain might inhibit inappropriate trapping of  
spermatozoa or promoting adhesion to the oviductal isthmus.  
-!- DOMAIN: The WFCD domains 2 and 3 may mediate covalent  
oligomerization (by similarity to human intestinal mucin MUC2).  
-!- SIMILARITY: Contains at least 2 MAM domains.  
-!- SIMILARITY: Contains 5 WFCD domains.  
-!- SIMILARITY: Contains 1 EGF-like domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF244982; AAF63342.2; -;  
HSSP; P56682; 1CCV.  
InterPro; IPR000742; EGF\_2.  
InterPro; IPR006209; EGF\_like.  
InterPro; IPR000998; MAM domain.  
InterPro; IPR002919; TIL\_Cysrich.  
InterPro; IPR003328; TILa\_Cysrich.  
InterPro; IPR001007; VWF\_C.  
InterPro; IPR001846; VWF\_D.  
Pfam; PF00008; EGF; 1.  
Pfam; PF00629; MAM; 2.  
Pfam; PF01826; TIL; 5.  
Pfam; PF02345; TILa; 5.  
Pfam; PF00094; vwd; 4.  
SMART; SM00137; MAM; 1.  
SMART; SM00214; VMC; 2.  
SMART; SM00216; VMD; 4.  
PROSITE; PS00022; EGF\_1; 1.  
PROSITE; PS01186; EGF\_2; 4.  
PROSITE; PS00026; EGF\_3; 1.  
PROSITE; PS00740; MAM\_1; FALSE\_NEG.  
PROSITE; PS00600; MAM\_2; 2.  
Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;  
Repeat.  
KW NON\_TER 1  
FT DOMAIN <1 2235  
FT TRANSMEM 2236 2256  
FT DOMAIN 2257 2282  
FT DOMAIN <1 147  
FT DOMAIN 150 315  
FT DOMAIN 315 498  
FT DOMAIN 499 610  
FT DOMAIN 611 995  
FT DOMAIN 996 1384  
FT DOMAIN 1385 1787  
FT DOMAIN 1788 2189  
FT DOMAIN 2189 2221  
FT DISULFID 2189 2200  
FT DISULFID 2194 2209  
FT DISULFID 2211 2220  
FT CARBOHYD 112 112  
FT CARBOHYD 272 272  
FT CARBOHYD 541 541  
FT CARBOHYD 569 569  
FT CARBOHYD 1141 1141  
FT CARBOHYD 1259 1259  
FT CARBOHYD 1270 1270  
FT CARBOHYD 1355 1355  
FT CARBOHYD 1467 1467  
FT CARBOHYD 1483 1483  
FT CARBOHYD 1662 1662  
FT CARBOHYD 1997 1997  
FT CARBOHYD 2178 2178  
FT CARBOHYD 2282 AA; 248290 MW; 380FA81093454892 CRC64;  
SQ SEQUENCE 78.68; Score 33; DB 1; Length 2282;  
Query Match



Pfam: PF00094; vwd; 3.  
 SMART: SM00041; CT; 1.  
 SMART: SM00231; FA58C; 2.  
 SMART: SM00192; LDLA; 1.  
 SMART: SM00216; VWD; 3.  
 PROSITE: PS01185; CTCK\_1; 1.  
 PROSITE: PS01225; CTCK\_2; 1.  
 PROSITE: PS01285; FA58C\_1; 2.  
 PROSITE: PS01286; FA58C\_2; 2.  
 PROSITE: PS01282; FA58C\_3; 2.  
 PROSITE: PS01208; VWFC\_1; FALSE NEG.  
 Lectin; Glycoprotein; Signal; Repeat; Cell adhesion.  
 SIGNAL: 1  
 CHAIN ? 3133  
 DOMAIN 29 131  
 DOMAIN 153 240  
 DOMAIN 248 613  
 DOMAIN 940 1095  
 DOMAIN 1116 1254  
 DOMAIN 1283 1356  
 DOMAIN 1620 1951  
 DOMAIN 1952 2315  
 DOMAIN 2320 2321  
 DOMAIN 2335 2361  
 DOMAIN 2435 2469  
 DOMAIN 2553 2622  
 DOMAIN 2842 2907  
 DOMAIN 2971 3076  
 DOMAIN 895 914  
 DOMAIN 1267 1270  
 DOMAIN 1425 1428  
 DOMAIN 1447 1450  
 DOMAIN 1474 1479  
 DOMAIN 2148 2153  
 DOMAIN 2156 2159  
 DOMAIN 2341 2344  
 DISULFID 940 1095  
 DISULFID 1116 1254  
 DISULFID 2981 3040  
 DISULFID 2991 3054  
 DISULFID 3004 3070  
 DISULFID 3020 3072  
 DISULFID ? 3075  
 CARBOHYD 151 151 (POTENTIAL)  
 CARBOHYD 237 237 (POTENTIAL)  
 CARBOHYD 564 564 (POTENTIAL)  
 CARBOHYD 1170 1170 (POTENTIAL)  
 CARBOHYD 1387 1387 (POTENTIAL)  
 CARBOHYD 1622 1622 (POTENTIAL)  
 CARBOHYD 1727 1727 (POTENTIAL)  
 CARBOHYD 1847 1847 (POTENTIAL)  
 CARBOHYD 1975 1975 (POTENTIAL)  
 CARBOHYD 1985 1985 (POTENTIAL)  
 CARBOHYD 2093 2093 (POTENTIAL)  
 CARBOHYD 2113 2113 (POTENTIAL)  
 CARBOHYD 2161 2161 (POTENTIAL)  
 CARBOHYD 2276 2276 (POTENTIAL)  
 CARBOHYD 2451 2451 (POTENTIAL)  
 CARBOHYD 2647 2647 (POTENTIAL)  
 CARBOHYD 2654 2654 (POTENTIAL)  
 CARBOHYD 2663 2663 (POTENTIAL)  
 CARBOHYD 2794 2794 (POTENTIAL)  
 CARBOHYD 2810 2810 (POTENTIAL)  
 CARBOHYD 2865 2865 (POTENTIAL)  
 CARBOHYD 2929 2929 (POTENTIAL)  
 CARBOHYD 2964 2964 (POTENTIAL)  
 CARBOHYD 3028 3028 (POTENTIAL)  
 VARIANT 1288 1288 R -> G.  
 VARIANT 1305 1305 T -> S.  
 SEQUENCE 3133 AA; 343350 MW; E5210D5D14A7B2B2 CRC64;  
 Query Match 78.6%; Score 33; DB 1; Length 3133;  
 Best Local Similarity 71.4%; Pred. No. 1.9e+02;

Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CVPLTSC 7  
 DB 203 CVPATQC 209

RESULT 11  
 MUC2 HUMAN  
 ID MUC2 HUMAN STANDARD; PRT; 5179 AA.  
 AC Q02817; Q14878;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Mucin 2 precursor (intestinal mucin 2).  
 GN MUC2 OR SMUC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=94132002; PubMed=8300571;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.  
 RT Identification of the amino terminus and overall sequence similarity  
 RT to prepro-von Willebrand factor.";  
 RL J. Biol. Chem. 269:2440-2446(1994).  
 RN [2]  
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=93016075; PubMed=1400449;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
 RA Kim Y.S.;  
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
 RT both upstream and downstream of its central repetitive region.";  
 RL J. Biol. Chem. 267:21375-21383(1992).  
 RN [3]  
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
 RX MEDLINE=91358717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.M., Kim Y.S.;  
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 RT and polymorphism.";  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and  
 CC other mucus membrane-containing organs. Thought to provide a  
 CC protective, lubricating barrier against particles and infectious  
 CC agents at mucosal surfaces.  
 CC -!- SUBUNIT: Multimeric.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
 CC BRONCHUS, CERVIX AND GALL BLADDER.  
 CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and  
 CC varies among different alleles.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 2 VWFC domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L21998; AAB95295.1; -;  
 CC EMBL; M74027; AAB59875.1; -;  
 CC EMBL; M94131; AAB59163.1; -;

TO BE USED FOR THE PURPOSES OF THE ACT.

```
CC -----
DR EMBL; S83181; AAB46908.1; -
DR EMBL; D38075; BAA22409.1; -
DR PIR; JC4790; JC4790.
DR InterPro; IPR002919; TIL_Cysrich.
KW Serine protease inhibitor; Signal.
FT SIGNAL 1 22
FT CHAIN 23 77
FT ACT_SITE 51 52
FT DISULFID 25 57
FT DISULFID 36 49
FT DISULFID 40 77
FT DISULFID 59 71
FT SEQUENCE 77 AA; 8492 MW; B9CFC085DDA10354 CRC64;
Query Match 76.2%; Score 32; DB 1; Length 77;
Best Local Similarity 57.1%; Pred. No. 7;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVPLTSC 7
Db 71 CVPISQC 77
RESULT 13
PEX2_RAT
ID PEX2_RAT STANDARD; PRT; 305 AA.
AC P24392; Q63733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisome assembly factor-1 (PAP-1) (Peroxin-2) (Peroxisomal membrane
protein 3).
GN PMP3 OR PEX2 OR PAF1 OR PMP35.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=91163637; PubMed=1750930;
RA Tsukamoto T., Miura S., Fujiki Y.;
RT "Restoration by a 35K membrane protein of peroxisome assembly in a
RL Peroxisome-deficient mammalian cell mutant.";
RL Nature 350:77-81(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=liver;
RX MEDLINE=94309666; PubMed=8035823;
RA Tsukamoto T., Shinozawa N., Fujiki Y.;
RT "Peroxisome assembly factor 1: nonsense mutation and deletion analysis.";
RL Mol. Cell. Biol. 14:5458-5465(1994).
CC 1- FUNCTION: Somewhat implicated in the biogenesis of peroxisomes.
CC 1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.
CC 1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; X57988; CAA41054.1; -
DR EMBL; D30616; BAA06306.1; -
DR EMBL; D30617; BAA06307.1; -
DR PIR; B53782; B53782.
DR InterPro; IPR006845; Pex2_Pex12.
DR InterPro; IPR001841; Znf_Ring.
DR Pfam; PF04757; Pex2_Pex12; 1.
```

```
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00589; ZF_RING_2; 1.
KW Transmembrane; Peroxisome; Zinc-finger.
FT TRANSMEM 140 159
FT TRANSMEM 195 213
FT ZN_FING 244 284
FT VARIANT 125 125 R -> P.
SQ SEQUENCE 305 AA; 34767 MW; 249058A5A5A8E854 CRC64;
Query Match 76.2%; Score 32; DB 1; Length 305;
Best Local Similarity 83.3%; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 CVPLTS 6
Db 224 CIPLTS 229
```

```
RESULT 14
PSPB_DICDI
ID PSPB_DICDI STANDARD; PRT; 379 AA.
AC P54704;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Prespore protein B precursor.
GN PSPB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=94326657; PubMed=8050366;
RA Powell-Coffman J.A., Firtel R.A.;
RT "Characterization of a novel Dictyostelium discoideum
RT prespore-specific gene, PspB, reveals conserved regulatory
sequences.";
RL Development 120:1601-1611(1994).
CC 1- SIMILARITY: Contains 3 prespore motif repeats.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; S72639; AAC0506.1; -
DR DictyBase; DDB0185060; pspB.
DR InterPro; IPR003645; FoLN.
DR SMART; SM00274; FoLN; 1.
KW Glycoprotein; Repeat; Sporulation; Signal.
FT SIGNAL 1 19
FT CHAIN 20 379
FT REPEAT 224 236 PRESPORE MOTIF 1.
FT REPEAT 289 301 PRESPORE MOTIF 2.
FT REPEAT 319 331 PRESPORE MOTIF 3.
SQ SEQUENCE 379 AA; 42719 MW; EA92A6ED69D05431 CRC64;
Query Match 76.2%; Score 32; DB 1; Length 379;
Best Local Similarity 57.1%; Pred. No. 35;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVPLTSC 7
Db 283 CIPKTC 289
```

RESULT 15

PM17\_BOVIN  
ID PM17\_BOVIN STANDARD; PRT; 491 AA.  
AC Q06154;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Melanocyte protein Pmel 17 (Retinal pigment epithelial-specific protein) (Fragment).  
DE SILV OR PMEL17 OR RPE1.  
GN Bos taurus (Bovine).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.  
OC NCBI\_TaxID=9913;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Retina;  
RC MEDLINE=93122163; PubMed=1478275;  
RA Kim R.Y.; Wistow G.J.;  
RT "The cDNA RPE1 and monoclonal antibody HMB-50 define gene products preferentially expressed in retinal pigment epithelium.";  
RL Exp. Eye Res. 55:657-662(1992).  
CC -!- FUNCTION: Could be a melanogenic enzyme (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: Retinal pigment epithelium.  
CC -!- SIMILARITY: BELONGS TO THE PMEL-17/NMB FAMILY.  
CC -!- SIMILARITY: Contains 1 PKD domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M81193; AAA30419.1; --  
DR PIR; A49179; A49179.  
DR InterPro; IPR000601; PKD.  
DR Pfam; PF00801; PKD; 1.  
DR SMART; SM00089; PKD; 1.  
DR PROSITE; PS50093; PKD; 1.  
KW Transmembrane; Glycoprotein; Melanin biosynthesis; Repeat.  
FT NON\_TER 1  
FT DOMAIN <1 423 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 424 444 POTENTIAL.  
FT DOMAIN 445 491 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 60 150 PKD.  
FT DOMAIN 148 256  
FT REPEAT 148 160  
FT REPEAT 161 173  
FT REPEAT 174 186  
FT REPEAT 187 199  
FT REPEAT 200 212  
FT REPEAT 213 225  
FT REPEAT 232 243  
FT REPEAT 244 256  
FT DOMAIN 304 394 CYS-RICH.  
FT CARBOHYD 269 269 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 396 396 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 491 AA; 51669 MW; 2BBFE5DFBD397D6D CRC64;  
  
Query Match 76.2%; Score 32; DB 1; Length 491;  
Best Local Similarity 83.3%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 VPLTSC 7  
Db 139 IPLTSC 144

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:46:09 ; Search time 15.8384 Seconds  
(without alignments)  
139,448 Million cell updates/sec

Title: US-09-761-636A-11

Perfect score: 42

Sequence: 1 CVPLTSC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: \*  
1: sp\_archea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvirus: \*  
16: sp\_bacteriaph: \*  
17: sp\_archeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	575	5 Q8IRB8	Q8IRB8 drosophila
2	39	92.9	2543	5 P90649	P90649 paramedium
3	38	90.5	127	6 Q9GMW9	Q9GMW9 macaca fasc
4	37	88.1	594	16 Q933X0	Q933X0 enterococcus
5	36	85.7	337	5 Q18464	Q18464 herdmamian m
6	36	85.7	980	2 O24780	O24780 clostridium
7	36	85.7	2533	5 P90589	P90589 paramedium
8	36	85.7	2533	5 Q27183	Q27183 paramedium
9	36	85.7	4998	11 Q8CG65	Q8CG65 mus musculus
10	35	83.3	193	13 Q90993	Q90993 gallus gall
11	35	83.3	379	16 Q82314	Q82314 chlamydomphi
12	35	83.3	389	6 Q97887	Q97887 bos taurus
13	35	83.3	462	11 Q9CXM8	Q9CXM8 mus musculus
14	35	83.3	565	4 Q8TF36	Q8TF36 homo sapien
15	35	83.3	685	4 Q8N8U9	Q8N8U9 homo sapien
16	35	83.3	685	11 Q8CJ69	Q8CJ69 mus musculus

17	35	83.3	685	11 Q7TNS7	Q7TNS7 mus musculus
18	35	83.3	745	4 Q8IYA6	Q8IYA6 homo sapien
19	35	83.3	1637	6 Q9XSV8	Q9XSV8 bos taurus
20	35	83.3	5146	6 Q8SPM4	Q8SPM4 bos taurus
21	34	81.0	912	6 Q8MIK0	Q8MIK0 crocuta cro
22	34	81.0	1506	13 P79927	P79927 xenopus lae
23	34	81.0	1736	4 Q8TES1	Q8TES1 homo sapien
24	34	81.0	2551	4 Q8WMQ8	Q8WMQ8 homo sapien
25	34	81.0	2551	4 Q8TUG9	Q8TUG9 homo sapien
26	34	81.0	2551	4 Q7ZSN9	Q7ZSN9 homo sapien
27	34	81.0	2843	4 Q9Y6R7	Q9Y6R7 homo sapien
28	33	78.6	55	10 Q38989	Q38989 arabidopsis
29	33	78.6	101	12 Q91F35	Q91F35 cvdia pomon
30	33	78.6	120	5 Q86BW2	Q86BW2 ascidia syd
31	33	78.6	143	16 Q9ACU2	Q9ACU2 streptomyce
32	33	78.6	191	6 Q28583	Q28583 ovis aries
33	33	78.6	196	11 Q9D226	Q9D226 mus musculus
34	33	78.6	207	16 Q25713	Q25713 helicobacte
35	33	78.6	207	16 Q9ZM78	Q9ZM78 helicobacte
36	33	78.6	223	11 Q62220	Q62220 mus musculus
37	33	78.6	230	11 Q64507	Q64507 mus musculus
38	33	78.6	245	10 Q8LKU4	Q8LKU4 sorghum bic
39	33	78.6	282	16 Q7VCX3	Q7VCX3 helicobacte
40	33	78.6	303	12 Q92585	Q92585 salmirine
41	33	78.6	303	12 Q40640	Q40640 salmirine
42	33	78.6	367	11 Q91VI9	Q91VI9 mus musculus
43	33	78.6	415	11 Q9JIE4	Q9JIE4 mus musculus
44	33	78.6	427	4 Q96Q28	Q96Q28 homo sapien
45	33	78.6	459	11 Q8BZJ8	Q8BZJ8 mus musculus

## ALIGNMENTS

## RESULT 1

Q8IRB8 PRELIMINARY; PRT; 575 AA.  
AC Q8IRB8  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE CG32260-PA.  
GN CG32260  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abirl J.F., Aghayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira J., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

QY 1 CVPLTSC 7  
Db 204 CLPLTSC 210

RESULT 2  
P90649 PRELIMINARY; PRT; 2543 AA.

ID P90649  
AC P90649 (T-EMBLrel. 03, Created)  
DT 01-MAY-1997 (T-EMBLrel. 03, Last sequence update)  
DE 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
DE 156D surface antigen.  
OS Paramacium primaurelia.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;  
OC Paramacium.  
ON NCBI\_TaxID=5886;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=156;  
RX MEDLINE=96313351; PubMed=8768434;  
RT Bourgoin-Guglielmetti F., Caron;  
RT "Molecular characterization of the D surface protein gene subfamily in  
RT Paramacium primaurelia.";  
RL J. Eukaryot. Microbiol. 43:303-314(1996).  
DR EMBL; X96616; CAA65436.1; -.  
DR PIR; T31687; T31687.  
DR InterPro; IPR002895; Paramacium\_SA.  
DR Pfam; PF01508; Paramacium\_SA; 20.  
DR SMART; SM00639; PSA; 25.  
SQ SEQUENCE 2543 AA; 267041 MW; 828EF797CB012902 CRC64;

Query Match 92.9%; Score 39; DB 5; Length 2543;  
Best Local Similarity 71.4%; Pred. No. 23;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
Db 2158 CIPITSC 2164

RESULT 3  
Q9GMW9 PRELIMINARY; PRT; 127 AA.

ID Q9GMW9  
AC Q9GMW9 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DE Hypothetical protein.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
OC Cercopitheidae; Macaca.  
ON NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Ceada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
RA Suzuki Y., Sugano S., Hashimoto K.;  
RT "Isolation of full-length cDNA clones from macaque brain cDNA  
RT libraries.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB047606; BAB12130.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 127 AA; 12961 MW; EE168525256992C0 CRC64;

Query Match 90.5%; Score 38; DB 6; Length 127;  
Best Local Similarity 85.7%; Pred. No. 2.8;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merklov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reiner K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Wang Z.Y., Wasserman D.A., Weinstock G.W., Wang A.H., Wang X.,  
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celisner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J.A., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.P., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA MacIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J.J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang X., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bergman C., Carlson J.W., Celisner S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB003480; AAN11589.1; -.  
DR FlyBase; FBgn0052260; CG32260.  
DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
DR GO; GO:0004295; F:trypsin activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR006604; CLIP.  
DR InterPro; IPR009003; Cys Ser trypsin.  
DR InterPro; IPR001254; Peptidase\_S1.  
DR InterPro; IPR001314; Peptidase\_S1A.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR SMART; SM00680; CLIP; 1.  
DR SMART; SM00020; TRYSP; 1.  
DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
SQ SEQUENCE 575 AA; 62063 MW; 9D2D66174BB99921 CRC64;

Query Match 92.9%; Score 39; DB 5; Length 575;  
Best Local Similarity 85.7%; Pred. No. 6.3;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

KW Hypothetical protein; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 337 UNKNOWN PROTEIN.
SQ SEQUENCE 337 AA; 35617 MW; 5CCA0924118D8FC6 CRC64;

Query Match 85.7%; Score 36; DB 5; Length 337;
Best Local Similarity 71.4%; Pred. No. 17;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 199 CVPISSC 205

RESULT 6
O24780 PRELIMINARY; PRT; 980 AA.
ID O24780
AC O24780;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Celodextrin phosphorylase.
GN CDP.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC27405;
RA Kawaguchi T., Ikeuchi Y., Tsutsumi N., Kan A., Sumitani J., Arai M.;
RT "Cloning, nucleotide sequence, and expression of the Clostridium
RT thermocellum celodextrin phosphorylase gene and its application to
RT synthesis of cellulase inhibitors.";
RL J. Ferment. Bioeng. 85:144-149(1998).
DR EMBL; AB006822; BAA22081.1; -.
DR FIRM; T00045; T00045.
DR InterPro; IPR008928; Glyco trans 6bp.
SQ SEQUENCE 980 AA; 111183 MW; 8A8C8EE3F5370831 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 980;
Best Local Similarity 57.1%; Pred. No. 42;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 550 CIPMTAC 556

RESULT 7
P90589 PRELIMINARY; PRT; 2533 AA.
ID P90589
AC P90589;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha-5iD immobilization antigen.
GN ALPHA-5iD.
OS Paramedium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramedium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=51;
RA Schwegmann K., Klein H., Schmidt H.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X96400; CAA65264.1; -.
DR FIRM; T28675; T28675.
DR InterPro; IPR002895; Paramedium SA.
DR Pfam; PF01508; Paramedium_SA; 21.
DR SMART; SM00633; PSA; 26.
SQ SEQUENCE 2533 AA; 264142 MW; EAED7F21E408C371 CRC64;

```

Query Match 85.7%; Score 36; DB 5; Length 2533;  
 Best Local Similarity 57.1%; Pred. No. 96;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
 Db 2149 CIPITNC 2155

RESULT 8  
 Q27183 ID Q27183 PRELIMINARY; PRT; 2533 AA.  
 AC Q27183;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Alpha-51D-immobilization antigen.  
 GN ALPHA-51D-GENE.  
 OS Paramesidium tetraurelia.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;  
 OC Paramesidium.  
 OX NCBI\_TaxID=5888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=51;  
 RA Schmidt H.J.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X85135; CAA59447.1; -;  
 DR PIR; T28674; T28674.  
 DR InterPro; IPR002895; Paramesidium SA.  
 DR Pfam; PF01508; Paramesidium\_SA; 22.  
 DR SMART; SM00639; PSA; 26.  
 SQ SEQUENCE 2533 AA; 263996 MW; 261BD09806BC344D CRC64;

Query Match 85.7%; Score 36; DB 5; Length 2533;  
 Best Local Similarity 57.1%; Pred. No. 96;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
 Db 2149 CIPITNC 2155

RESULT 9  
 Q8CG65 ID Q8CG65 PRELIMINARY; PRT; 4998 AA.  
 AC Q8CG65;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE SCO-spondin.  
 DE SCO-SPONDIN.  
 GN Mus musculus (Mouse).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR; TISSUE=Brain;  
 RA Goncalves N., Simon-Chazottes D., Creveaux I., Meinzel A.,  
 RA Guenet J.-L., Meinzel R.;  
 RT "Characterization, spatio-temporal expression and chromosomal assignment of mouse SCO-spondin."  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ491857; CAD42654.1; -;  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005507; F:copper ion binding; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0005179; F:hormone activity; IEA.  
 DR GO; GO:0007155; P:cell adhesion; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR000923; BlueCu\_1.

DR InterPro; IPR006207; Cys\_knot\_C.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR000421; FA58\_C.  
 DR InterPro; IPR001545; Gly\_hormoneB.  
 DR InterPro; IPR002172; LDL\_receptor\_A.  
 DR InterPro; IPR002919; TIL\_Cysrich.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR000885; TSP1.  
 DR InterPro; IPR006552; VC\_out.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF00754; F5\_F8\_type\_C; 1.  
 DR Pfam; PF00057; ldl\_recept\_a; 10.  
 DR Pfam; PF01826; TIL; 10.  
 DR Pfam; PF00090; tsp\_1; 25.  
 DR Pfam; PF00093; wvc7\_1.  
 DR Pfam; PF00094; wvd; 3.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PRINTS; PR01705; TSP1REPEAT.  
 DR SMART; SM00192; LDLa; 10.  
 DR SMART; SM00209; TSP1; 25.  
 DR SMART; SM00214; VMC; 4.  
 DR SMART; SM00215; VMC\_out; 9.  
 DR SMART; SM00216; VMD; 2.  
 DR PROSITE; PS00196; COPPER\_BLUE; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00222; FA58C\_3; 1.  
 DR PROSITE; PS00261; GLYCO\_HORMONE\_BETA\_1; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 8.  
 DR PROSITE; PS00068; LDLRA\_2; 10.  
 DR PROSITE; PS00092; TSP1; 25.  
 DR PROSITE; PS01208; VWF\_C\_1; 1.  
 DR PROSITE; PS0184; VWF\_C\_2; 2.  
 SQ SEQUENCE 4998 AA; 535019 MW; DA2ABAB8DA47DF225 CRC64;

Query Match 85.7%; Score 36; DB 11; Length 4998;  
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
 Db 4570 CVPISCC 4576

RESULT 10  
 Q90993 ID Q90993 PRELIMINARY; PRT; 193 AA.  
 AC Q90993;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Pp60c-scr protein.  
 DE PP60C-SCR.  
 GN Gallus gallus (Chicken).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pectoralis muscle;  
 RX MEDLINE=90318371; PubMed=2115117;  
 RA Dorai T., Wang L.-H.;  
 RT "An alternative non-tyrosine protein kinase product of the c-src gene in chicken skeletal muscle";  
 RL Mol. Cell. Biol. 10:4068-4079(1990).  
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; M57290; AAA49078.1; -;  
 DR PIR; A00630; TVCHS.  
 DR HSSP; P00523; LSRL.  
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro; IPR000980; SH2.



```

DR InterPro; IPR001452; SH3.
DR Pfam; PF00017; SH2; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000093; SH2; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 193 AA; 21180 MW; 9D1B54C33B02D98 CRC64;

Query Match 83.3%; Score 35; DB 13; Length 193;
Best Local Similarity 71.4%; Pred. No. 17;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
Db 185 CIPLPSC 191

RESULT 11
Q82314 PRELIMINARY; PRT; 379 AA.
ID Q82314
AC Q82314
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN CCA00426.
OS Chlamydomophila caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI_TaxID=83557;
RN [1]
RP STRAIN=GPIC;
RC MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapfel E., Khouri H., Federova N.B., Carty H.A.,
RA Mayhew L.A., Haft D.H., Peterson J.J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoli P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae."
RL Nucleic Acids Res. 31:2134-2147(2003).
DR EMBL; AF016995; AAP05172.1; -.
DR TIGR; CCA00426; -.
DR InterPro; IPR002589; Alpp.
DR SMART; SM00506; Alpp; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 379 AA; 40634 MW; 62879277DBADD55D CRC64;

Query Match 83.3%; Score 35; DB 16; Length 379;
Best Local Similarity 71.4%; Pred. No. 30;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
Db 13 CIPLTSC 19

RESULT 12
O97887 PRELIMINARY; PRT; 389 AA.
ID O97887
AC O97887
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Reissner's fiber glycoprotein I (fragment).
OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Subcommissural organ;
RX MEDLINE=98228246; PubMed=9560467;
RA Nualart F., Hein S., Yulis C.R., Zarraga A.M., Araya A.,
RA Rodriguez E.M.;
RT "Partial sequencing of Reissner's fiber glycoprotein I (RF-Gly I).";
RL Cell Tissue Res. 292:239-250(1998).
DR EMBL; AF078930; AADI7695.1; -.
DR HSRF; P56682; ICCV.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR InterPro; IPR001545; Gly_hormoneB.
DR InterPro; IPR003919; TIL_Cysrich.
DR InterPro; IPR008884; TSPI.
DR Pfam; PF01826; TIL; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS00261; GLYCO_HORMONE_BETA_1; 1.
DR PROSITE; PS50092; TSPI_1; 1.
FT NON_TER
SQ SEQUENCE 389 AA; 40720 MW; 48571E1F16F9F1B6 CRC64;

Query Match 83.3%; Score 35; DB 6; Length 389;
Best Local Similarity 71.4%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
Db 311 CVPVSSC 317

RESULT 13
Q9CXM8 PRELIMINARY; PRT; 462 AA.
ID Q9CXM8
AC Q9CXM8
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 3110056H04Rik protein.
GN CV2 OR 3110056H04Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustingich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Wyszewski-Boris A., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).

```

EMBL; AK014221; BAB29213.1; --  
 DR HSP; P00214; 1FTC.  
 DR MGD; MGI:1920480; CV2.  
 DR InterPro; IPR002919; TIL Cysrich.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; vwc; 2.  
 DR Pfam; PF00094; vwd; 1.  
 DR SMART; SM00214; VMC; 2.  
 DR SMART; SM00216; VMD; 1.  
 DR PROSITE; PS01208; VWF\_C\_1; 1.  
 DR PROSITE; PS01208; VWF\_C\_2; 1.  
 DR PROSITE; PS01208; VWF\_C\_3; 1.  
 SQ SEQUENCE 462 AA; 51641 MW; 586AF92A52D7A4B6 CRC64;

Query Match 83.3%; Score 35; DB 11; Length 462;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVELTSC 7  
 |||:|  
 Db 113 CVPVSSC 119

RESULT 14  
 Q8TF36 PRELIMINARY; PRT; 565 AA.  
 ID Q8TF36  
 AC Q8TF36  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein KIAA1965 (Fragment).  
 GN KIAA1965.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX MEDLINE=21842142; PubMed=11853319;  
 RA Nagase T., Kikuno R., Ohara O.;  
 RT Prediction of the coding sequences of unidentified human genes. XXII.  
 RT The complete sequences of 50 new cDNA clones which code for large  
 RT proteins.;  
 RL DNA Res. 8:319-327(2001).  
 DR EMBL; AB075845; BAB85551.1; --  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR002919; TIL Cysrich.  
 DR InterPro; IPR000294; VitK\_dep\_GLA.  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; vwc; 3.  
 DR Pfam; PF00094; vwd; 1.  
 DR SMART; SM00214; VMC; 3.  
 DR SMART; SM00216; VMD; 1.  
 DR PROSITE; PS00011; GLU CARBOXYLATION; 1.  
 DR PROSITE; PS01208; VWF\_C\_1; 2.  
 DR PROSITE; PS01208; VWF\_C\_2; 2.  
 KW Hypothetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 565 AA; 62934 MW; 1F7A15206F2B9D45 CRC64;

Query Match 83.3%; Score 35; DB 4; Length 565;  
 Best Local Similarity 57.1%; Pred. No. 42;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVELTSC 7  
 |||:|  
 Db 216 CIPVSSC 222

# RESULT 15

Q8N8U9 PRELIMINARY; PRT; 685 AA.  
 ID Q8N8U9  
 AC Q8N8U9  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ38831 (Crossveinless-2).  
 DE Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX MEDLINE=9606;  
 SQ SEQUENCE FROM N.A.  
 RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
 RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
 RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,  
 RA Osuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,  
 RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RA "NEDO human cDNA sequencing project";  
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [1]  
 SQ SEQUENCE FROM N.A.  
 RA Binnerts M.E., Wen X., Cante-Barrett K., Bright J., Chen H.-T.,  
 RA Asundi V., Sattari P., Tang T., Boyle B., Funk W., Rupp F.;  
 RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK096150; BAC04712.1; --  
 DR EMBL; AY324650; AAP89012.1; --  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR InterPro; IPR002919; TIL Cysrich.  
 DR InterPro; IPR000294; VitK\_dep\_GLA.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; vwc; 3.  
 DR Pfam; PF00094; vwd; 1.  
 DR SMART; SM00214; VMC; 5.  
 DR SMART; SM00216; VMD; 1.  
 DR PROSITE; PS00011; GLU CARBOXYLATION; 1.  
 DR PROSITE; PS01208; VWF\_C\_1; 3.  
 DR PROSITE; PS01208; VWF\_C\_2; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 685 AA; 75996 MW; A5048AADEBF04AB0 CRC64;

Query Match 83.3%; Score 35; DB 4; Length 685;  
 Best Local Similarity 57.1%; Pred. No. 50;  
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVELTSC 7  
 |||:|  
 Db 336 CIPVSSC 342

Search completed: September 5, 2004, 10:00:03  
 Job time : 17.8384 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:37:49 ; Search time 22.9798 Seconds  
(without alignments)  
86.068 Million cell updates/sec

Title: US-09-761-636A-11

Perfect score: 42

Sequence: 1 CVPLTSC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Genesep 29Jan04:\*

- 1: Genesep1980s:\*
- 2: Genesep1990s:\*
- 3: Genesep2000s:\*
- 4: Genesep2001s:\*
- 5: Genesep2002s:\*
- 6: Genesep2003s:\*
- 7: Genesep2003bs:\*
- 8: Genesep2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	7	AAU04530	AAU04530 VEGF base
2	38	90.5	50	AAU62562	AAU62562 Propionib
3	38	90.5	50	ABM59081	ABM59081 Propionib
4	37	90.5	74	ABG13170	ABG13170 Novel hum
5	37	88.1	64	ABM16042	ABM16042 Human ner
6	36	85.7	91	AAU64239	AAU64239 Propionib
7	36	85.7	91	ABM60758	ABM60758 Propionib
8	36	85.7	4561	ABG30203	ABG30203 Novel hum
9	36	85.7	9222	ABG21064	ABG21064 Novel hum
10	35	83.3	61	AAU96565	AAU96565 Human rep
11	35	83.3	61	ABM96569	ABM96569 Human tes
12	35	83.3	65	AAU18859	AAU18859 Novel pro
13	35	83.3	65	AAU94953	AAU94953 Human rep
14	35	83.3	80	AAU49012	AAU49012 Propionib
15	35	83.3	80	ABM45531	ABM45531 Propionib
16	35	83.3	270	AAU99932	AAU99932 Human pol
17	35	83.3	270	ABG65083	ABG65083 Human alb
18	35	83.3	281	ADG07963	ADG07963 Novel pro
19	35	83.3	445	AAE07062	AAE07062 Human gen
20	35	83.3	445	ABG65086	ABG65086 Human alb
21	35	83.3	464	AAE07119	AAE07119 Human gen
22	35	83.3	627	AAU99293	AAU99293 Human cho
23	35	83.3	685	AAU99292	AAU99292 Human cho
24	35	83.3	685	ABU11857	ABU11857 Human sec
25	35	83.3	685	ADE07923	ADE07923 Novel pro

## ALIGNMENTS

## RESULT 1

AAU04530  
ID AAU04530 standard; peptide; 7 AA.

XX AC AAU04530;

DT 26-SEP-2001 (first entry)

XX DE VEGF based monocyclic peptide 8.

XX Human; VEGF; vascular endothelial growth factor; angiogenesis;  
neovascularisation; lymphangiogenesis; psoriasis; tumour;  
diabetes induced neovascular sequelae; rheumatoid arthritis;  
diabetic retinopathy; chronic inflammation; cyclic.  
XX Synthetic.

XX Key Location/Qualifiers

FT Disulfide-bond 1..7 /note= "This bond cyclises the peptide"

XX WO200152875-A1.

XX 26-JUL-2001.

XX 18-JAN-2001; 2001WO-US001533.

XX 18-JAN-2000; 2000US-0176293P.

XX 16-MAY-2000; 2000US-0204590P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Achen MG, Hughes RA, Stacker S, Cendron A;

XX WPI; 2001-442248/47.

XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
from an exposed loop of a growth factor protein by oxidizing the cysteine  
residues.

XX Claim 49; Page 32; 102pp; English.

XX The sequence represents a monomeric monocyclic peptide of the invention,  
whose 3-dimensional structure is modelled on the exposed loop of human  
VEGFD (vascular endothelial growth factor). The invention relates to a  
method of producing a monomeric monocyclic peptide by a measuring beta-  
beta carbon separation distances on opposite antiparallel strands of a

CC peptide loop fragment from an exposed loop of a growth factor protein and  
CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
CC to cyclisation are used to interfere with angiogenesis,  
CC neovascularisation or lymphangiogenesis in a mammal with a condition  
CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
CC hemangioma, vascularised malignant or benign tumour, post-recovery  
CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
CC trauma, substance-induced neovascularisation of the liver, excessive  
CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
CC infection. The peptides are also used to modulate vascular permeability  
CC in a mammal (the mammal has a condition characterised by fluid  
CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
CC or brain. The peptides are used to image blood vessels and lymphatic  
CC vasculature. The monomeric and bicyclic peptides are used to interfere  
CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
CC are also used in combination with an anti-inflammatory agent, to treat a  
CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
CC diabetic retinopathy  
XX

SQ Sequence 7 AA;  
Query Match 100.0%; Score 42; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
| | | | |  
DB 1 CVPLTSC 7

RESULT 2  
AAU62562  
ID AAU62562 standard; protein; 50 AA.  
XX AC AAU62562;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #23458.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.  
XX OS Propionibacterium acnes.  
XX PN WO200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US012865.  
XX PR 21-APR-2000; 2000US-0199047P.  
XX PR 02-JUN-2000; 2000US-0208941P.  
XX PR 07-JUL-2000; 2000US-0216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
XX DR N-PSDB; AAS5962.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris.

PS Example 1; SEQ ID NO 23757; 1069pp; English.  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 50 AA;  
Query Match 90.5%; Score 38; DB 4; Length 50;  
Best Local Similarity 85.7%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
| | | | |  
DB 22 CAPLTSC 28

RESULT 3  
ABM59081  
ID ABM59081 standard; protein; 50 AA.  
XX AC ABM59081;  
XX DT 20-OCT-2003 (first entry)  
XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #23757.  
XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
XX immunostimulant; immune response; vaccine.  
XX OS Propionibacterium acnes.  
XX PN WO2003033515-A1.  
XX PD 24-APR-2003.  
XX PF 11-OCT-2002; 2002WO-US032727.  
XX PR 15-OCT-2001; 2001US-00978825.  
XX PA (CORI-) CORIXA CORP.  
XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
XX Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
XX Barth B, Vallieve-Douglass J;  
XX WPI; 2003-381789/36.  
XX DR N-PSDB; ACF64556.  
XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
XX polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
XX or for stimulating an immune response specific for a P. acnes protein.  
XX Example 1; SEQ ID NO 23757; 1481pp; English.  
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC

CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide; a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 50 AA;

Query Match 90.5%; Score 38; DB 6; Length 50;  
 Best Local Similarity 85.7%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPILTSC 7  
 |||||  
 Db 22 CAPLTSC 28

RESULT 4  
 ABG13170  
 ID ABG13170 standard; protein; 74 AA.

AC ABG13170;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #13161.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS77357.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX Claim 20; SEQ ID NO 43529; 103pp; English.  
 ES

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 74 AA;

Query Match 90.5%; Score 38; DB 4; Length 74;  
 Best Local Similarity 85.7%; Pred. No. 53;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPILTSC 7  
 |||||  
 Db 29 CVPILSC 35

RESULT 5  
 ABB16042

ID ABB16042 standard; protein; 64 AA.

AC ABB16042;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polypeptide SEQ ID NO 4699.

XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antischlicking; antianemic; antiarthritis; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

OS Homo sapiens.

PN WO200159063-A2.

PD 16-AUG-2001.

PF 17-JAN-2001; 2001WO-US001334.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

us-09-761-636a-11.open.rag

Sun Sep 5 10:36:18 2004

PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 16-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 28-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0242221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-541565/60.  
XX N-PSDB; ABA12368.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
PT  
XX  
PS Claim 11; SEQ ID NO 4699; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,

CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 64 AA;

Query Match 88.1%; Score 37; DB 4; Length 64;  
 Best Local Similarity 85.7%; Pred. No. 68;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVPLTSC 7  
 Db 28 CXPLTSC 34

RESULT 6  
 AAU64239  
 ID AAU64239 standard; protein; 91 AA.  
 AC AAU64239;  
 XX  
 DT 27-FEB-2002 (first entry)  
 DE Propionibacterium acnes immunogenic protein #25135.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US012865.  
 XX  
 PR 21-APR-2000; 2000US-0199047P.  
 PR 02-JUN-2000; 2000US-0208841P.  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59641.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 XX  
 PS Example 1; SEQ ID NO 25434; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 91 AA;

Query Match 85.7%; Score 36; DB 4; Length 91;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
 Db 75 CVPLTSC 81

RESULT 7  
 ABM60758  
 ID ABM60758 standard; protein; 91 AA.  
 AC ABM60758;  
 XX  
 DT 20-OCT-2003 (first entry)  
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #25434.  
 XX  
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO2003033515-A1.  
 XX  
 PD 24-APR-2003.  
 XX  
 PF 11-OCT-2002; 2002WO-US032727.  
 XX  
 PR 15-OCT-2001; 2001US-00978825.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MU, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallieue-Douglas J;  
 XX  
 DR WPI; 2003-381789/36.  
 DR N-PSDB; ACF64570.  
 XX  
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 PS Example 1; SEQ ID NO 25434; 1481pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a

Query Match	85.7%;	Score 36;	DB 4;	Length 4561;
Best Local Similarity	71.4%;	Pred. No.	4.4e+03;	
Conservation	9.1%	Mismatches	0;	Gaps 0;

Qy	1	CVPLTSC	7
		:	
Dh	2931	CVPISSC	2937

RESULT 9  
ABG21064

ABG21064  
ID ARG21064 standard; protein; 9222 AA.

AC ABG21064;

18-FEB-2002 (first entry)

DE Novel human diagnostic protein #21055.

XX  
KW . Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX  
XX . Food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

AA  
PN  
W0200175067-A2.

PD 11-OCT-2001.

30-MAR-2001: 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

XX

XX

XX

DR N-PSDB; AAS85251.

PT New isolated polynucleotide and encoded polypeptide, used for  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

It is not responsible for

PS Claim 20; SEQ ID NO 51423; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (I) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders



CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 9222 AA;

Query Match 85.7%; Score 36; DB 4; Length 9222;

Best Local Similarity 71.4%; Pred. No. 8.2e+03;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 3655 CVPISSC 3661

RESULT 10

AAW96565

ID AAW96565 standard; protein; 61 AA.

XX AC AAW96565;

XX 21-NOV-2001 (first entry)

XX Human reproductive system related antigen SEQ ID NO: 5223.

KW Human; reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy.

XX OS Homo sapiens.

XX WO200153320-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US0001339.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 26-JUL-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225266P.

XX 14-AUG-2000; 2000US-0225267P.

XX 14-AUG-2000; 2000US-0225268P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229345P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.

PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.

PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.

PR 14-SEP-2000; 2000US-0232397P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232398P.

PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.

PR 14-SEP-2000; 2000US-0232401P.

PR 14-SEP-2000; 2000US-0233063P.

PR 14-SEP-2000; 2000US-0233064P.

PR 21-SEP-2000; 2000US-0234223P.

PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.

PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.

PR 27-SEP-2000; 2000US-0235834P.

PR 29-SEP-2000; 2000US-0235836P.

PR 29-SEP-2000; 2000US-0236327P.

PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.

PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.

PR 02-OCT-2000; 2000US-0237039P.

PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239335P.

PR 13-OCT-2000; 2000US-0239337P.

PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.

PR 20-OCT-2000; 2000US-0241221P.

PR 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.

PR 20-OCT-2000; 2000US-0241809P.

PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246474P.

PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246475P.

PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.

PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.

PR 08-NOV-2000; 2000US-0246525P.

PR 08-NOV-2000; 2000US-0246526P.

PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.

PR 08-NOV-2000; 2000US-0246532P.

PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.

PR 08-NOV-2000; 2000US-0246611P.

reproductive system disorder; urinary system disorder; gene therapy;  
cardiovascular disorder; respiratory disorder; neurological disorder;  
gastrointestinal disease; infection; cytostatic.

Homo sapiens.

WO200155317-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US001329.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180628P.

24-FEB-2000; 2000US-0184664P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205151P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214896P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

07-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

11-JUL-2000; 2000US-0217496P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0224519P.

14-AUG-2000; 2000US-0225213P.

14-AUG-2000; 2000US-0225214P.

14-AUG-2000; 2000US-0225266P.

14-AUG-2000; 2000US-0225267P.

14-AUG-2000; 2000US-0225268P.

14-AUG-2000; 2000US-0225270P.

14-AUG-2000; 2000US-0225447P.

14-AUG-2000; 2000US-0225757P.

14-AUG-2000; 2000US-0225758P.

14-AUG-2000; 2000US-0225759P.

18-AUG-2000; 2000US-0226279P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226686P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

30-AUG-2000; 2000US-0228924P.

08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249212P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249244P.  
17-NOV-2000; 2000US-0249245P.  
17-NOV-2000; 2000US-0249246P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
01-DEC-2000; 2000US-0250391P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
05-DEC-2000; 2000US-0256719P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-02559678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

N-PSDB; AAL02535.

Isolated nucleic acid molecule encoding a reproductive system antigen is  
used in preventing, treating or ameliorating a medical condition.

Claim 11; SEQ ID NO 5223; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a  
number of human reproductive system related antigens. These can be used  
in the prevention and treatment of reproductive system disorders,  
including cancer. The present sequence is a protein of the invention

Sequence 61 AA;

Query Match 83.3%; Score 35; DB 4; Length 61;

Best Local Similarity 85.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTSC 7

Db 40 CVPLGSC 46

RESULT 11

ABB96569

ID ABB96569 standard; protein; 61 AA.

XX AC ABB96569;

XX AC ABB96569;

DT 21-JUN-2002 (first entry)

XX DE Human testicular antigen SEQ ID NO: 1953.

XX DE Human; testicular antigen; testes; cancer; metastasis; immune disorder;

KW

PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244826P.  
PR 08-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483232/52.  
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful for preventing, diagnosing and/or treating testicular cancer.  
XX Claim 11; SEQ ID NO 1953; 766pp; English.  
XX The present invention provides the protein and coding sequences of 973 human testicular antigens, and fragments of their genomic sequences. The sequences can be used in the treatment of cardiovascular, urinary system, reproductive system, immune, respiratory, neurological and gastrointestinal disorders, infections, and particularly cancer, especially testicular cancers. The present sequence is a protein of the invention  
XX SQ Sequence 61 AA;  
Query Match 83.3%; Score 35; DB 4; Length 61;  
Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CVPLTSC 7  
Db 40 CVPLGSC 46  
|||||  
RESULT 12  
AAU18859  
ID AAU18859 standard; protein; 65 AA.  
XX AAU18859;  
XX 21-NOV-2001 (first entry)  
XX Novel prostate gland antigen, Seq ID No 158.  
XX Human; neutropic; neuroprotective; cytostatic; antiparkinsonian;  
KW antinaemic; dermatological; immunosuppressive; antiinflammatory;  
KW antiarthritic; antirheumatic; virucide; hepatotropic; nephrotropic;  
KW osteopathic; prostate gland; prostatitis; adenocarcinoma; hair loss;  
KW hyperplasia; carcinoma; adenocarcinoma; benign prostatic hypertrophy;  
KW reproductive system disorder; autoimmune disorder; skin aging;  
KW systemic lupus erythematosus; rheumatoid arthritis; urinary system;  
KW blood-related disorder; hyperproliferative disorder; cardiovascular;  
KW neurological disorder; endocrine disorder; inflammatory disorder;  
KW liver disorder; wound healing; food preservative.  
XX Homo sapiens.  
XX WO200155447-A1.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001330.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218230P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228944P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-476223/51.  
XX N-PSDB; AAS30300.  
XX Novel isolated prostate gland related polypeptide useful for diagnosis  
PT and treatment of disorders of prostate such as prostatodystonia,  
PT prostatosis, prostatitis, benign prostatic hypertrophy and malacoplakia.  
XX Claim 1; SEQ ID NO 158; 512pp; English.  
PS The invention relates to novel isolated prostate gland related nucleic  
XX acids (I) and polypeptides (II). (I) and (II) are useful for diagnosis,  
CC prognosis, prevention, and/or treatment of diseases and/or disorders of  
CC the prostate such as acute non-bacterial prostatitis, chronic non-  
CC bacterial prostatitis, acute bacterial prostatitis, prostatodystonia,  
CC prostatosis, granulomatous prostatitis, malacoplakia, benign prostatic

CC hypertrophy or hyperplasia, and prostate neoplastic disorders, including  
 CC adenocarcinomas, transitional cell carcinomas, ductal carcinomas, and  
 CC squamous cell carcinomas. (I), (II) and antibody to (II) are useful for  
 CC diagnosing and treating reproductive system disorders (Paget's disease),  
 CC autoimmune disorders (systemic lupus erythematosus, rheumatoid  
 CC arthritis), blood-related disorders (sickle cell anaemia),  
 CC hyperproliferative disorders, urinary system disorders  
 CC (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory  
 CC disorders, musculoskeletal system disorders, neural activity and  
 CC neurological disorders (Alzheimer's disease and Parkinson's disease),  
 CC endocrine disorders (Addison's disease), gastrointestinal disorders  
 CC (inflammatory disorders), liver disorders (biliary liver cirrhosis),  
 CC pancreatic and gall bladder disorders, disorders of the large intestine,  
 CC developmental and inherited disorders, diseases at the cellular level,  
 CC and wound healing and epithelial cell proliferation. (I) or (II) is  
 CC useful to prevent skin aging, for preventing hair loss, to maintain

Query Match 83.3%; Score 35; DB 4; Length 65;  
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
 Db 41 CLPLSSC 47

## RESULT 13

AA094953  
 ID AA094953 standard; protein; 65 AA.

AC AA094953;

XX 21-NOV-2001 (first entry)

XX Human reproductive system related antigen SEQ ID NO: 3611.

DE Human; reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy.

XX Homo sapiens.

XX WC200155320-A2.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US001339.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 14-AUG-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 13-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.



XX Propionibacterium acnes predicted ORF-encoded polypeptide #10207.  
DE  
XX  
KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO2003033515-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 11-OCT-2002; 2002WO-US032727.  
XX  
PR 15-OCT-2001; 2001US-00978825.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieve-Douglas J;  
XX  
DR WPI; 2003-381789/36.  
DR N-PSDB; ACF64472.  
XX  
XX  
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
PS  
FS Example 1; SEQ ID NO 10207; 1481pp; English.  
XX  
XX  
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC polynucleotide of the invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 80 AA;

Query Match 83.3%; Score 35; DB 6; Length 80;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
| | | | |  
Db 37 CDPLTSC 43

Search completed: September 5, 2004, 09:55:17  
Job time : 25.9798 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 10:00:15 ; Search time 19.6566 Seconds  
(without alignments)  
112.199 Million cell updates/sec

Title: US-09-761-636A-11

Perfect score: 42

Sequence: 1 CVPLTSC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	7	9	US-09-761-636A-11
2	38	90.5	118	12	US-10-424-599-248060
3	37	88.1	253	12	US-10-424-599-199668
4	36	85.7	75	16	US-10-437-963-118478
5	35	83.3	61	10	US-09-764-891-5223
6	35	83.3	65	10	US-09-764-891-3611
7	35	83.3	196	14	US-10-120-018-5
8	35	83.3	270	11	US-09-833-245-1832
9	35	83.3	270	14	US-10-091-458-48
10	35	83.3	270	15	US-10-191-254-48
11	35	83.3	445	11	US-09-833-245-1835
12	35	83.3	462	12	US-10-016-248-58
13	35	83.3	462	14	US-10-120-018-16
14	35	83.3	464	14	US-10-120-018-17
15	35	83.3	627	16	US-10-451-417-4

16	83.3	646	14	US-10-120-018-12	Sequence 12, Appl
17	83.3	685	12	US-10-263-929-116	Sequence 116, App
18	83.3	685	14	US-10-120-018-9	Sequence 9, Appli
19	83.3	685	15	US-10-108-260A-3546	Sequence 3546, Ap
20	83.3	685	16	US-10-451-417-2	Sequence 2, Appli
21	83.3	867	12	US-10-112-944-424	Sequence 424, App
22	83.3	867	11	US-09-833-245-1243	Sequence 1243, Ap
23	81.0	109	11	US-10-029-386-29056	Sequence 29056, A
24	81.0	221	12	US-10-112-944-465	Sequence 465, App
25	81.0	351	12	US-10-112-944-906	Sequence 906, App
26	81.0	1653	14	US-10-133-172-20	Sequence 20, Appl
27	81.0	2420	15	US-10-028-248A-4	Sequence 4, Appli
28	81.0	2420	15	US-10-107-782-4	Sequence 4, Appli
29	81.0	2675	15	US-10-028-248A-2	Sequence 2, Appli
30	81.0	2675	15	US-10-107-782-2	Sequence 2, Appli
31	81.0	5405	9	US-09-922-217-1116	Sequence 1116, Ap
32	81.0	5405	13	US-10-025-380-1116	Sequence 1116, Ap
33	78.6	39	12	US-10-424-599-189382	Sequence 189382,
34	78.6	55	12	US-10-296-115-748	Sequence 748, App
35	78.6	107	12	US-10-424-599-158802	Sequence 158802,
36	78.6	125	16	US-10-437-963-149232	Sequence 149232,
37	78.6	130	12	US-10-424-599-215673	Sequence 215673,
38	78.6	159	12	US-10-425-114-41558	Sequence 41558, A
39	78.6	207	9	US-09-881-752A-46	Sequence 46, Appl
40	78.6	207	12	US-10-335-977-8827	Sequence 8827, Ap
41	78.6	218	12	US-10-335-977-8828	Sequence 8828, Ap
42	78.6	370	16	US-10-437-963-104490	Sequence 104490,
43	78.6	375	14	US-10-029-386-32138	Sequence 32138, A
44	78.6	401	14	US-10-029-386-33904	Sequence 33904, A
45	78.6	414	14	US-10-043-487-219	Sequence 219, App

#### ALIGNMENTS

#### RESULT 1

US-09-761-636A-11  
; Sequence 11, Application US/09761636A  
; Patent No. US20020065218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 11  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-761-636A-11

Query Match 100.0%; Score 42; DB 9; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7

Db 1 CVPLTSC 7

#### RESULT 2

US-10-424-599-248060  
; Sequence 248060, Application US/10424599  
; Publication No. US20040031072A1

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 248060
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_66028C.1.pep
US-10-424-599-248060

Query Match          90.5%; Score 38; DB 12; Length 118;
Best Local Similarity 85.7%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 29 CTPLTSC 35

RESULT 3
US-10-424-599-199668
; Sequence 199668, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 199668
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_22324C.1.pep
US-10-424-599-199668

Query Match          88.1%; Score 37; DB 12; Length 253;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 109 CVPLASC 115

RESULT 4
US-10-437-963-118478
; Sequence 118478, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
```

```
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 118478
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_21786C.1.pep
US-10-437-963-118478

Query Match          85.7%; Score 36; DB 16; Length 75;
Best Local Similarity 71.4%; Pred. No. 66;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 53 CIPLTHC 59

RESULT 5
US-09-764-891-5223
; Sequence 5223, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5223
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (60)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-5223

Query Match          83.3%; Score 35; DB 10; Length 61;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 40 CVPLGSC 46

RESULT 6
US-09-764-891-3611
; Sequence 3611, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3611
; LENGTH: 65
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-3611

Query Match      83.3%; Score 35; DB 10; Length 65;
Best Local Similarity 71.4%; Pred. No. 86;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
DB 41 CLPLSSC 47

RESULT 7
US-10-120-018-5
; Sequence 5, Application US/10120018
; Publication No. US20030194708A1
; GENERAL INFORMATION:
; APPLICANT: Binnerts et al
; TITLE OF INVENTION: Human Homolog of Crossveinless Materials and Methods
; FILE REFERENCE: 28110/37995
; CURRENT APPLICATION NUMBER: US/10/120,018
; CURRENT FILING DATE: 2002-04-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-120-018-5

Query Match      83.3%; Score 35; DB 14; Length 196;
Best Local Similarity 57.1%; Pred. No. 2.3e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
DB 50 CIPSSC 56

RESULT 8
US-09-833-245-1832
; Sequence 1832, Application US/09833245
; Publication No. US2004010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1832
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (113)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (118)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

; LOCATION: (268)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-1832

Query Match      83.3%; Score 35; DB 11; Length 270;
Best Local Similarity 57.1%; Pred. No. 3.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7
DB 161 CIPSSC 167

RESULT 9
US-10-091-458-48
; Sequence 48, Application US/10091458
; Publication No. US20030068627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT206C1
; CURRENT APPLICATION NUMBER: US/10/091,458
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 09/764,900
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-03-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/236,369
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/224,519
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,964
; PRIOR FILING DATE: 2000-07-26
```

; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214

; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/235,836  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/249,218  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,208  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,213  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,212  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,207  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,245  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,244  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,217  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,211  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,215  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,264  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,214  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,297  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/232,400  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/231,242  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,081  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,080  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,414  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,244  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/233,064  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/233,063  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,397  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,399  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,401  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/241,808  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,826  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,786  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,221  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,475  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/231,243  
; PRIOR FILING DATE: 2000-09-08

Query Match 83.3%; Score 35; DB 14; Length 270;  
Best Local Similarity 57.1%; Pred. No. 3.1e+02;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
QY      1  CVPLTSC 7
Db      161  CIPSSC 167

RESULT 10
US-10-191-254-48
; Sequence 48, Application US/10191254
; Publication No. US2004005650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ06C2
; CURRENT APPLICATION NUMBER: US/10/191,254
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: 10/091,458
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 09/764,900
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (113)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (268)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-191-254-48

Query Match      83.3%; Score 35; DB 15; Length 270;
Best Local Similarity 57.1%; Pred. No. 3.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1  CVPLTSC 7
Db      161  CIPSSC 167

RESULT 11
US-09-833-245-1835
; Sequence 1835, Application US/09833245
```

```
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1835
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (147)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (288)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (332)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (443)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-1835

Query Match      83.3%; Score 35; DB 11; Length 445;
Best Local Similarity 57.1%; Pred. No. 4.9e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1  CVPLTSC 7
Db      336  CIPSSC 342

RESULT 12
US-10-016-248-58
; Sequence 58, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
```

```
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-016-248-58

Query Match      83.3%; Score 35; DB 12; Length 462;
Best Local Similarity 71.4%; Pred. No. 5.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 113 CVPVSSC 119

RESULT 13
US-10-120-018-16
; Sequence 16, Application US/10120018
; Publication No. US20030194708A1
; GENERAL INFORMATION:
; APPLICANT: Binner et al
; TITLE OF INVENTION: Human Homolog of Crossveinless Materials and Methods
; FILE REFERENCE: 28110/37995
; CURRENT APPLICATION NUMBER: US/10/120,018
; CURRENT FILING DATE: 2002-04-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-120-018-16

Query Match      83.3%; Score 35; DB 14; Length 462;
Best Local Similarity 71.4%; Pred. No. 5.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 113 CVPVSSC 119

RESULT 14
US-10-120-018-17
; Sequence 17, Application US/10120018
; Publication No. US20030194708A1
; GENERAL INFORMATION:
; APPLICANT: Binner et al
; TITLE OF INVENTION: Human Homolog of Crossveinless Materials and Methods
; FILE REFERENCE: 28110/37995
; CURRENT APPLICATION NUMBER: US/10/120,018
; CURRENT FILING DATE: 2002-04-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (166)..(166)
; OTHER INFORMATION: wherein "X" is unknown or other
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (307)..(307)
; OTHER INFORMATION: wherein "X" is unknown or other
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (312)..(312)
; OTHER INFORMATION: wherein "X" is unknown or other
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (351)..(351)
; OTHER INFORMATION: wherein "X" is unknown or other

; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (462)..(462)
; OTHER INFORMATION: wherein "X" is unknown or other
US-10-120-018-17

Query Match      83.3%; Score 35; DB 14; Length 464;
Best Local Similarity 57.1%; Pred. No. 5.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 355 CIPISSC 361

RESULT 15
US-10-451-417-4
; Sequence 4, Application US/10451417
; Publication No. US20040110921A1
; GENERAL INFORMATION:
; APPLICANT: ORITA, Satoshi et al
; TITLE OF INVENTION: A NOVEL POLYPEPTIDE AND ITS DNA
; FILE REFERENCE: 0032-0276P
; CURRENT APPLICATION NUMBER: US/10/451,417
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: JP 401584/2000
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-451-417-4

Query Match      83.3%; Score 35; DB 16; Length 627;
Best Local Similarity 57.1%; Pred. No. 6.7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7
Db 336 CIPISSC 342

Search completed: September 5, 2004, 10:29:21
Job time : 20.6566 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:55:30 ; Search time 6.22222 Seconds  
(without alignments)  
58.079 Million cell updates/sec

Title: US-09-761-636A-11  
Perfect score: 42  
Sequence: 1 CVPLTSC 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	616	US-09-134-000C-6380	Sequence 6380, Ap
2	34	81.0	394	US-09-252-991A-29358	Sequence 29358, A
3	34	81.0	2476	US-08-276-967-2	Sequence 2, Appl
4	34	81.0	2594	US-08-718-388-7	Sequence 7, Appl
5	34	81.0	5405	US-08-718-388-9	Sequence 9, Appl
6	33	78.6	280	US-08-969-106-11	Sequence 11, Appl
7	33	78.6	280	US-09-338-125-11	Sequence 11, Appl
8	33	78.6	376	US-08-875-972-2	Sequence 2, Appl
9	32	76.2	68	US-09-904-615-92	Sequence 92, Appl
10	32	76.2	142	US-09-252-991A-22994	Sequence 22994, A
11	32	76.2	150	US-09-252-991A-33012	Sequence 33012, A
12	32	76.2	382	US-09-252-991A-29595	Sequence 29595, A
13	32	76.2	661	US-08-417-174-27	Sequence 27, Appl
14	32	76.2	661	US-08-417-174-121	Sequence 121, Appl
15	32	76.2	661	US-08-231-565A-27	Sequence 27, Appl
16	32	76.2	661	US-09-007-961-27	Sequence 27, Appl
17	32	76.2	661	US-09-267-439-121	Sequence 121, Appl
18	32	76.2	661	US-08-388-852B-2	Sequence 2, Appl
19	32	76.2	661	US-09-073-138-27	Sequence 27, Appl
20	32	76.2	661	US-09-073-138-121	Sequence 121, Appl
21	32	76.2	661	US-07-891-942G-6	Sequence 6, Appl
22	32	76.2	661	US-09-134-000C-6451	Sequence 6451, Ap
23	32	76.2	710	US-08-626-710-21	Sequence 21, Appl
24	31	73.8	7	US-08-862-855-21	Sequence 21, Appl
25	31	73.8	7	US-09-226-985-21	Sequence 21, Appl
26	31	73.8	7	US-09-227-906-21	Sequence 21, Appl
27	31	73.8	7		

Sequence 35, Appl  
Sequence 36, Appl  
Sequence 37, Appl  
Sequence 38, Appl  
Sequence 23, Appl  
Sequence 94, Appl  
Sequence 12, Appl  
Sequence 3, Appl  
Sequence 11, Appl  
Sequence 17, Appl  
Sequence 10, Appl  
Sequence 11, Appl  
Patent No. 5194596  
Patent No. 5219739  
Patent No. 5194596  
Patent No. 5219739  
Sequence 15, Appl

28 31 73.8 12 3 US-08-742-243-35  
29 31 73.8 12 3 US-08-742-243-36  
30 31 73.8 12 3 US-08-742-243-37  
31 31 73.8 12 3 US-08-742-243-38  
32 31 73.8 17 3 US-08-807-992B-23  
33 31 73.8 45 4 US-09-716-129-94  
34 31 73.8 65 4 US-09-244-583-12  
35 31 73.8 109 3 US-08-691-794-3  
36 31 73.8 110 4 US-09-392-932-11  
37 31 73.8 110 4 US-09-574-708A-11  
38 31 73.8 110 4 US-09-822-270-17  
39 31 73.8 110 4 US-09-392-931-10  
40 31 73.8 110 4 US-09-392-931-11  
41 31 73.8 120 6 5194596-9  
42 31 73.8 120 6 5219739-9  
43 31 73.8 121 6 5194596-19  
44 31 73.8 121 6 5219739-20  
45 31 73.8 136 4 US-09-037-983C-15

## ALIGNMENTS

RESULT 1  
US-09-134-000C-6380  
; Sequence 6380, Application US/09134000C  
; Patent No. 8617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 6380  
; LENGTH: 616  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-6380

Query Match 88.1%; Score 37; DB 4; Length 616;  
Best Local Similarity 71.4%; Pred. No. 83;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTSC 7  
|:|:|:|  
Db 507 CLPMTSC 513

RESULT 2  
US-09-252-991A-29358  
; Sequence 29358, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29358  
; LENGTH: 394  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29358

Query Match 81.0%; Score 34; DB 4; Length 394;  
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 169 CMPCTSC 175

## RESULT 3

US-08-276-967-2

; Sequence 2, Application US/08276967

; Patent No. 5851817

; GENERAL INFORMATION:

; APPLICANT: Hardy, Daniel M.

; APPLICANT: Garbers, David L.

; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of

; TITLE OF INVENTION: Sperm

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White &amp; Durkee

; STREET: P. O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210-4433

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/276.967

; FILING DATE: Submitted Herewith

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Kitchell, Barbara S.

; REGISTRATION NUMBER: 33,928

; REFERENCE/DOCKET NUMBER: UTSD:418\KIT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 713-787-1400

; TELEFAX: 713-789-2679

; TELEX: 79-0924

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2476 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-276-967-2

Query Match

Best Local Similarity 81.0%; Score 34; DB 2; Length 2476;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 1505 CVPLSQC 1511

## RESULT 4

US-08-718-388-7

; Sequence 7, Application US/08718388

; Patent No. 6271362

; GENERAL INFORMATION:

; APPLICANT: MORIKAWA, MINORU

; APPLICANT: HARADA, NAOKI

; TITLE OF INVENTION: GENE ENCODING IGG Fc REGION-BINDING

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

; STREET: PO BOX 747

; CITY: FALLS CHURCH

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/718,388

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: MURPHY JR, GERALD M

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 0230-111

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2594 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-718-388-7

Query Match

Best Local Similarity 81.0%; Score 34; DB 3; Length 2594;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7

Db 1183 CLPLASC 1189

## RESULT 5

US-08-718-388-9

; Sequence 9, Application US/08718388

; Patent No. 6271362

; GENERAL INFORMATION:

; APPLICANT: MORIKAWA, MINORU

; APPLICANT: HARADA, NAOKI

; TITLE OF INVENTION: GENE ENCODING IGG Fc REGION-BINDING

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

; STREET: PO BOX 747

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/718,388

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: MURPHY JR, GERALD M

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 0230-111

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050



; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5405 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-718-388-9

Query Match 81.0%; Score 34; DB 3; Length 5405;  
Best Local Similarity 71.4%; Pred. No. 2e+03;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPILTSC 7  
|:|:|:|  
Db 1183 CLPLASC 1189

RESULT 6  
US-08-969-106-11  
; Sequence 11, Application US/08969106  
; Patent No. 5986055  
; GENERAL INFORMATION:  
; APPLICANT: Yang, M.  
; APPLICANT: Nandabalan, K.  
; APPLICANT: Schulz, V.  
; TITLE OF INVENTION: CDK2 INTERACTIONS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/969,106  
; FILING DATE: 13-NOV-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7934-057  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 280 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-969-106-11

Query Match 78.6%; Score 33; DB 2; Length 280;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPILTSC 7  
|:|:|:|  
Db 227 CVPATEC 233

RESULT 7  
US-09-338-125-11  
; Sequence 11, Application US/09338125  
; Patent No. 6521412

; GENERAL INFORMATION:  
; APPLICANT: Yang, M.  
; APPLICANT: Nandabalan, K.  
; APPLICANT: Schulz, V.  
; TITLE OF INVENTION: CDK2 INTERACTIONS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/338,125  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/969,106  
; FILING DATE: 13-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7934-057  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 280 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-09-338-125-11

Query Match 78.6%; Score 33; DB 4; Length 280;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPILTSC 7  
|:|:|:|  
Db 227 CVPATEC 233

RESULT 8  
US-08-875-972-2  
; Sequence 2, Application US/08875972  
; Patent No. 5985564  
; GENERAL INFORMATION:  
; APPLICANT: Huntington Potter and Jinhue Li  
; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173-4799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/875,972  
FILING DATE: 08-AUG-97  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/002,448  
FILING DATE: 16-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan Esq., Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: HU95-03PA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240  
TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 376 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-875-972-2

Query Match 78.6%; Score 33; DB 2; Length 376;  
Best Local Similarity 71.4%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
Db 26 CVPVTL 32

RESULT 9  
US-09-904-615-92  
Sequence 92, Application US/09904615  
Patent No. 6566325  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 49 Human Secreted Proteins  
FILE REFERENCE: P2032P1  
CURRENT APPLICATION NUMBER: US/09/904,615  
CURRENT FILING DATE: 2001-07-16  
PRIOR APPLICATION NUMBER: 09/511,554  
PRIOR FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: 60/097,917  
PRIOR FILING DATE: 1998-08-25  
PRIOR APPLICATION NUMBER: 60/098,634  
PRIOR FILING DATE: 1998-08-31  
NUMBER OF SEQ ID NOS: 170  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 92  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (68)  
OTHER INFORMATION: Xaa equals stop translation  
US-09-904-615-92

Query Match 76.2%; Score 32; DB 4; Length 68;  
Best Local Similarity 83.3%; Pred. No. 72;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPLTSC 7  
Db 38 IPLTSC 43

RESULT 10  
US-09-252-991A-22994  
Sequence 22994, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22994  
LENGTH: 142  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22994

Query Match 76.2%; Score 32; DB 4; Length 142;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
Db 19 CCPTSC 25

RESULT 11  
US-09-252-991A-33012  
Sequence 33012, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 33012  
LENGTH: 150  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-33012

Query Match 76.2%; Score 32; DB 4; Length 150;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
Db 90 CMPPTSC 96

RESULT 12  
US-09-252-991A-29595  
Sequence 29595, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 29595

; LENGTH: 382  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29595

Query Match 76.2%; Score 32; DB 4; Length 382;  
Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CVPLTSC 7  
|||||  
Db 76 CVPLQOC 82

RESULT 13  
US-08-417-174-27  
; Sequence 27, Application US/08417174  
; Patent No. 5844075  
; GENERAL INFORMATION:  
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
; APPLICANT: STEVEN A.  
; TITLE OF INVENTION: MELANOMA ANTIGENS AND  
; THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
; METHODS  
; NUMBER OF SEQUENCES: 126  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417,174  
; FILING DATE: 05-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,565  
; FILING DATE: 22-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CAROL M. GRUPPI  
; REGISTRATION NUMBER: 37,341  
; REFERENCE/DOCKET NUMBER: 2026-4124US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 661  
; TYPE: amino acid  
; STRANDEDNESS: Unknown  
; TOPOLOGY: Unknown  
; MOLECULE TYPE: Protein  
US-08-417-174-27

Query Match 76.2%; Score 32; DB 2; Length 661;  
Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPLTSC 7  
|||||  
Db 296 IFLTSC 301

RESULT 14  
US-08-417-174-121  
; Sequence 121, Application US/08417174  
; Patent No. 5844075  
; GENERAL INFORMATION:  
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
; APPLICANT: STEVEN A.  
; TITLE OF INVENTION: MELANOMA ANTIGENS AND  
; THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
; METHODS  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
US-08-417-174-121

Query Match 76.2%; Score 32; DB 2; Length 661;  
Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPLTSC 7  
|||||  
Db 296 IFLTSC 301

RESULT 15  
US-08-231-565A-27  
; Sequence 27, Application US/08231565A  
; Patent No. 5874560  
; GENERAL INFORMATION:  
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
; APPLICANT: STEVEN A.  
; TITLE OF INVENTION: MELANOMA ANTIGENS AND  
; THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
; METHODS  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
US-08-231-565A-27

; Patent No. 5844075  
; GENERAL INFORMATION:  
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
; APPLICANT: STEVEN A.  
; TITLE OF INVENTION: MELANOMA ANTIGENS AND  
; THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
; METHODS  
; NUMBER OF SEQUENCES: 126  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/417,174  
; FILING DATE: 05-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,565  
; FILING DATE: 22-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CAROL M. GRUPPI  
; REGISTRATION NUMBER: 37,341  
; REFERENCE/DOCKET NUMBER: 2026-4124US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 121:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 661  
; TYPE: amino acid  
; STRANDEDNESS: Unknown  
; TOPOLOGY: Unknown  
; MOLECULE TYPE: Protein  
US-08-417-174-121

Query Match 76.2%; Score 32; DB 2; Length 661;  
Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPLTSC 7  
|||||  
Db 296 IFLTSC 301

RESULT 15  
US-08-231-565A-27  
; Sequence 27, Application US/08231565A  
; Patent No. 5874560  
; GENERAL INFORMATION:  
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
; APPLICANT: STEVEN A.  
; TITLE OF INVENTION: MELANOMA ANTIGENS AND  
; THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
; METHODS  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
US-08-231-565A-27

Query Match 76.2%; Score 32; DB 2; Length 661;  
Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 VPLTSC 7  
|||||  
Db 296 IFLTSC 301

RESULT 15  
US-08-231-565A-27  
; Sequence 27, Application US/08231565A  
; Patent No. 5874560  
; GENERAL INFORMATION:  
; APPLICANT: KAWAKAMI, YUTAKA; ROSENBERG,  
; APPLICANT: STEVEN A.  
; TITLE OF INVENTION: MELANOMA ANTIGENS AND  
; THEIR USE IN DIAGNOSTIC AND THERAPEUTIC  
; METHODS  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
US-08-231-565A-27

```

;
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/231,565A
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 661
; TYPE: amino acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Unknown
; MOLECULE TYPE: Protein
;
US-08-231-565A-27

```

```

Query Match          76.2%; Score 32; DB 2; Length 661;
Best Local Similarity 83.3%; Pred. No. 6.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 VPLTSC 7
       :|||||
Db      296 IPLTSC 301

```

```

Search completed: September 5, 2004, 10:21:59
Job time : 6.2222 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:47:29 ; Search time 4.60606 Seconds  
(without alignments)  
125.302 Million cell updates/sec

Title: US-09-761-636A-12

Perfect score: 38

Sequence: 1 CVPLTC 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: pirl:\*

2: pirl2:\*

3: pirl3:\*

4: pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	817	2 T25674	hypothetical prote
2	37	97.4	293	2 T27621	hypothetical prote
3	35	92.1	1627	2 S65464	pregnancy-associat
4	34	89.5	367	2 T20914	hypothetical prote
5	33	86.8	332	2 I48933	adenosine receptor
6	33	86.8	415	2 A54126	endothelin recepto
7	33	86.8	467	2 AC2390	transposase all467
8	33	86.8	637	2 T08050	probable ethylene
9	33	86.8	650	2 A34498	glycoprotein anti
10	33	86.8	1236	1 A53970	1-phosphatidylinos
11	32	84.2	108	2 S12982	hypothetical 12.2K
12	32	84.2	137	2 T26433	hypothetical prote
13	32	84.2	156	2 C81364	probable periplasm
14	32	84.2	252	2 B82659	two-component syst
15	32	84.2	301	2 T40593	cytoplasmic dynein
16	32	84.2	314	2 B43652	ferredoxin [2Fe-2S
17	32	84.2	345	1 NBBO	apolipoprotein H p
18	32	84.2	398	2 A35281	integumentary muc
19	32	84.2	425	2 T40610	probable clathrin
20	32	84.2	518	2 AD1930	ammonium transport
21	32	84.2	972	2 A03663	glycoprotein GP330
22	32	84.2	1034	2 JC5598	mucin - rat
23	32	84.2	4660	2 T42737	gp330 protein prec
24	31	81.6	40	2 E45495	beta-defensin-5 -
25	31	81.6	133	2 T29606	hypothetical prote
26	31	81.6	224	2 T27063	hypothetical prote
27	31	81.6	326	2 C69341	cobalt transport A
28	31	81.6	414	2 T50019	hypothetical prote
29	31	81.6	434	2 I50132	vasotocin receptor

30 31 81.6 471 2 AC3622  
31 31 81.6 588 2 T24121  
32 31 81.6 718 2 JE0188  
33 31 81.6 822 2 T01622  
34 31 81.6 869 1 JC4858  
35 31 81.6 1042 2 T26644  
36 31 81.6 1313 2 T29027  
37 31 81.6 1451 2 T30821  
38 31 81.6 2098 2 T13166  
39 31 81.6 2946 2 T00867  
40 30 78.9 40 2 G45495  
41 30 78.9 40 2 I45495  
42 30 78.9 42 2 C45495  
43 30 78.9 42 2 F45495  
44 30 78.9 64 2 A47438  
45 30 78.9 64 2 A56128

mannose-1-phosphat  
hypothetical prote  
RES4-25 protein -  
probable salt-indu  
VLDL receptor prec  
hypothetical prote  
hypothetical prote  
p-glycoprotein - S  
rough deal protein  
hypothetical prote  
beta-defensin-7 -  
beta-defensin-9 -  
beta-defensin-3 -  
beta-defensin-6 -  
airway epithelial  
lingual antimicrob

#### ALIGNMENTS

##### RESULT 1

T25674

hypothetical protein F08D12.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25674

R;Je, T.; Waterston, R.

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid F08D12.

A:Reference number: Z20068

A:Accession: T25674

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-817 <LET>

A:Cross-references: EMBL:U08040; PIDN:AA837929.1; GSPDB:GN00020; CESP:F08D12.6

A:Experimental source: strain Bristol N2; clone F08D12

C:Genetics:

A:Gene: CESP:F08D12.6

A:Map position: 2

A:Introns: 27/2; 329/3; 407/1; 477/3; 538/3; 622/2

Query Match 100.0%; Score 38; DB 2; Length 817;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6

Db 479 CVPLTC 484

##### RESULT 2

T27621

hypothetical protein ZC504.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T27621

R;Kershaw, J.

submitted to the EMBL Data Library, July 1995

A:Reference number: Z20394

A:Accession: T27621

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-293 <WIL>

A:Cross-references: EMBL:Z50029; PIDN:CAA90343.1; GSPDB:GN00028; CESP:ZC504.5

A:Experimental source: clone ZC504

C:Genetics:

A:Gene: CESP:ZC504.5

A:Map position: X

A:Introns: 40/3; 140/3; 173/3; 255/3

Query Match 97.4%; Score 37; DB 2; Length 293;  
Best Local Similarity 83.3%; Pred. No. 12;

R; LLOYD, C.  
submitted to the EMBL Data Library. March 1997

24

D6 254 CVPLAC 259

D6 254 CVPLAC 259

## RESULT 7

AC2390  
 transposase all4675 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AC2390  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoro, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC2390  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-467 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA076374.1; PID:gl7133812; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all4675

Query Match 86.8%; Score 33; DB 2; Length 467;  
 Best Local Similarity 83.3%; Pred. No. 94;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 |||||  
 Db 61 CCPLTC 66

## RESULT 8

T08050  
 probable ethylene receptor - muskmelon  
 C:Species: Cucumis melo (muskmelon)  
 C>Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999  
 C:Accession: T08050  
 R:Sato-Nara, K.; Yuhashi, K.; Higashi, K.; Hosoya, K.; Kubota, M.; Ezura, H.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: Expression of an ERS homolog gene of melon (Cucumis melo L. reticulatus)  
 A:Reference number: Z16317  
 A:Accession: T08050  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-637 <SAT>  
 A:Cross-references: EMBL:AF037368; NID:q2707333; PIDN:AAC99477.1; PID:g2707334

Query Match 86.8%; Score 33; DB 2; Length 637;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 |||||  
 Db 226 CVPTTC 231

## RESULT 9

A34498  
 Glycoprotein antigen Bm86 precursor - southern cattle tick  
 C:Species: Boophilus microplus (southern cattle tick)  
 C>Date: 22-Jun-1990 #sequence\_revision 22-Jun-1990 #text\_change 21-Jul-2003  
 C:Accession: A34498  
 R:Rand, K.N.; Moore, T.; Srikantha, A.; Spring, K.; Tellam, R.; Willadsen, P.; Cobon, C.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 9657-9661, 1989  
 A:Title: Cloning and expression of a protective antigen from the cattle tick Boophilus m.  
 A:Reference number: A34498; MUID:90099323; PMID:2690068  
 A:Accession: A34498  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-650 <RAN>  
 A:Cross-references: GB:M20321; NID:gl161667; PID:gl161668  
 A:Keywords: glycoprotein  
 F:209-246/Domain: EGF homology <EGF>

Query Match 86.8%; Score 33; DB 2; Length 650;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 |||||  
 Db 147 CVPTTC 152

## RESULT 10

A53970  
 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC 3.1.4.11) gamma-D - fruit  
 C:Species: Drosophila melanogaster  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 17-Mar-2000  
 C:Accession: A53970  
 R:Emori, Y.; Sugaya, R.; Akimaru, H.; Higashijima, S.; Shishido, E.; Saigo, K.; Homma, Y.  
 J. Biol. Chem. 269, 19474-19479, 1994  
 A:Title: Drosophila phospholipase C-gamma expressed predominantly in blastoderm cells at  
 A:Reference number: A53970; MUID:94308233; PMID:8034716  
 A:Accession: A53970  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1236 <EMO>  
 A:Cross-references: GB:D29806

C:Genetics:  
 A:Gene: plc-gammad  
 A:Cross-references: FlyBase:FBgn0003416  
 C:Superfamily: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase II; 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase domain Y hon

Query Match 86.8%; Score 33; DB 1; Length 1236;  
 Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 |||||  
 Db 1196 CVPLTC 1201

## RESULT 11

S12982  
 hypothetical 12.2K protein (68K signal recognition particle region) - dog  
 C:Species: Canis lupus familiaris (dog)  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999  
 C:Accession: S12982  
 R:Herz, J.; Flint, N.; Stanley, K.; Frank, R.; Dobberstein, B.  
 FEBS Lett. 276, 103-107, 1990  
 A:Title: The 68 kDa protein of signal recognition particle contains a glycine-rich region  
 A:Reference number: S12981; MUID:91092392; PMID:1702390  
 A:Accession: S12982  
 A:Molecule type: mRNA  
 A:Residues: 1-108 <HER>  
 A:Cross-references: GB:X53744; NID:g931; PIDN:CAA37774.1; PID:g933  
 C:Superfamily: dog hypothetical 12.2K protein (68K signal recognition particle region)

Query Match 84.2%; Score 32; DB 2; Length 108;  
 Best Local Similarity 66.7%; Pred. No. 43;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 |||||  
 Db 20 CVPMS 25

## RESULT 12

T26433

hypothetical protein Y106G6H.16 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T26433

R:McMurray, A.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z2014

A/Accession: T26433

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-137 <WIL>

A/Cross-references: EMBL:AL032631; PIDN:CAA21579.1; GSPDB:GN00019; CESP:Y106G6H.16

A/Experimental source: clone Y106G6H

C/Genetics:

A/Gene: CESP:Y106G6H.16

A/Map position: 1

A/Introns: 26/3; 61/3; 89/3

Query Match	84.2%	Score 32;	DB 2;	Length 137;
Best Local Similarity	83.3%	Pred. No. 52;		
Matches	5;	Conservative 0;	Mismatches 1;	Indels 0;
Gaps	0;			

Qy	1	CVPLTC 6
Db	45	CVPLTC 50

#### RESULT 13

C81364

probable periplasmic protein Cj0908 [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni

C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C/Accession: C81364

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

F.; Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

Nature 403, 665-668, 2000

A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A/Reference number: A81250; MUID:20150912; PMID:10688204

A/Accession: C81364

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-156 <PAR>

A/Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAB73165.1; PID:G696834

A/Experimental source: serotype O2, strain NCTC 11168

C/Genetics:

A/Gene: Cj0908

Query Match	84.2%	Score 32;	DB 2;	Length 156;
Best Local Similarity	83.3%	Pred. No. 58;		
Matches	5;	Conservative 0;	Mismatches 1;	Indels 0;
Gaps	0;			

Qy	1	CVPLTC 6
Db	36	CVPLTC 41

#### RESULT 14

B82659

two-component system, regulatory protein XF1626 [imported] - Xylella fastidiosa (strain

C:Species: Xylella fastidiosa

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2003

C/Accession: B82659

R:Anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A/Reference number: A82515; MUID:20365717; PMID:10910347

A/Note: for a complete list of authors see reference number A59328 below

A/Accession: B82659

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-252 <SIM>

A/Cross-references: GB:AE003989; GB:AE003849; NID:G9106666; PIDN:AAF84435.1; GSPDB:GN001

A/Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, F. as-Neto, E.; Docena, C.; El-Dorry, H.; Facinca, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A/Authors: Ferreira, V.C.A.; Ferrio, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Promm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, C.Y.

A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawas

A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;

A/Reference number: A59328

C/Contents: annotation

C/Genetics:

A/Gene: XF1626

C/Superfamily: response regulator with LytR DNA-binding domain, ALGR/VlirR/ComE type; r

Query Match	84.2%	Score 32;	DB 2;	Length 252;
Best Local Similarity	83.3%	Pred. No. 86;		
Matches	5;	Conservative 0;	Mismatches 1;	Indels 0;
Gaps	0;			

Qy	1	CVPLTC 6
Db	234	CVPLTC 239

Qy	1	CVPLTC 6
Db	234	CVPLTC 239

Qy	1	CVPLTC 6
Db	234	CVPLTC 239

#### RESULT 15

T40593

cytoplasmic dynein intermediate chain - fission yeast (Schizosaccharomyces pombe) (frag

C:Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C/Accession: T40593

R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, January 1999

A/Reference number: Z21938

A/Accession: T40593

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-301 <SSE>

A/Cross-references: EMBL:AL035216; PIDN:CAA22821.1; GSPDB:GN000067; SPDB:SPBC646.17c

A/Experimental source: strain 972h-; cosmid c646

C/Genetics:

A/Gene: SPDB:SPBC646.17c

A/Map position: 2

Query Match	84.2%	Score 32;	DB 2;	Length 301;
Best Local Similarity	66.7%	Pred. No. 1e+02;		
Matches	4;	Conservative 1;	Mismatches 1;	Indels 0;
Gaps	0;			

Qy	1	CVPLTC 6
Db	101	CIPATC 106

Search completed: September 5, 2004, 10:01:22

Job time : 5.60606 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:38:39 ; Search time 2.72727 Seconds  
(without alignments)  
114.554 Million cell updates/sec

Title: US-09-761-636A-12

Perfect score: 38

Sequence: 1 CVPLTC 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	94.7	3564	1	CSM1_MOUSE
2	36	94.7	3565	1	CSM1_HUMAN
3	35	92.1	530	1	PDP2_RAT
4	35	92.1	1627	1	PAPA_HUMAN
5	35	92.1	3487	1	CSM2_HUMAN
6	33	86.8	265	1	TPIS_SYNEL
7	33	86.8	332	1	AA2B_MOUSE
8	33	86.8	650	1	EM86_BOOMI
9	33	86.8	1587	1	LMG3_HUMAN
10	32	84.2	314	1	VANB_PSES9
11	32	84.2	345	1	APOH_BOVIN
12	32	84.2	345	1	APOH_PANTR
13	32	84.2	370	1	DCAM_TRYBB
14	32	84.2	398	1	MOB1_XENLA
15	32	84.2	940	1	CHRD_BRARE
16	32	84.2	1581	1	LMG3_MOUSE
17	32	84.2	4655	1	LRP2_HUMAN
18	32	84.2	4660	1	LRP2_RAT
19	31	81.6	64	1	BD05_BOVIN
20	31	81.6	96	1	K121_HUMAN
21	31	81.6	146	1	K122_HUMAN
22	31	81.6	167	1	OB_SMICR
23	31	81.6	188	1	V060_F0WV
24	31	81.6	434	1	AVT_CATCO
25	31	81.6	810	1	NEL1_HUMAN
26	31	81.6	2796	1	CSM3_MOUSE
27	31	81.6	3670	1	CSM3_HUMAN
28	30	78.9	38	1	BD08_BOVIN
29	30	78.9	40	1	BD07_BOVIN
30	30	78.9	40	1	BD07_BOVIN
31	30	78.9	42	1	BD06_BOVIN
32	30	78.9	55	1	BD09_BOVIN
33	30	78.9	57	1	BD03_BOVIN

34	30	78.9	64	1	EAP_BOVIN
35	30	78.9	64	1	LAP_BOVIN
36	30	78.9	64	1	TAP_BOVIN
37	30	78.9	194	1	KRUB_HUMAN
38	30	78.9	223	1	YK83_CABEL
39	30	78.9	313	1	DHSL_HUMAN
40	30	78.9	313	1	DHSL_MOUSE
41	30	78.9	426	1	ETIR_RAT
42	30	78.9	427	1	ETIR_BOVIN
43	30	78.9	427	1	ETIR_HUMAN
44	30	78.9	427	1	ETIR_MOUSE
45	30	78.9	427	1	ETIR_PIG

## RESULT 1

CSM1\_MOUSE  
ID\_CSM1\_MOUSE STANDARD; PRT; 3564 AA.  
AC Q923L3; Q8BUV1; Q8BYQ3;  
DT 15-MAR-2004 (Rel. 43, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE CUB and eushi multiple domains protein 1 precursor.  
GN CSM1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC STRAIN=C57BL/6;  
RX MEDLINE=21365705; PubMed=11472063;  
RA Sun P.C., Uppaluri R., Schmidt A.P., Bashia M.E., Quant E.C.,  
RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;  
RT "Transcript map of the 8p23 putative tumor suppressor region.";  
RL Genomics 75:17-25(2001).  
[2]  
RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 3250-3564 FROM N.A. (ISOFORM 2).  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Hypothalamus;  
RX MEDLINE=22354693; PubMed=12466851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavalon K., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).

-!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).



CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC (1)  
 CC SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 CC MEDLINE=21365705; PubMed=11472063;  
 CC Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,  
 CC Sunwoo J.B., Gollin S.M., Scholnick S.B.;  
 CC "Transcript map of the 8p23 putative tumor suppressor region.";   
 CC Genomics 75:17-25(2001).  
 CC (2)  
 CC SEQUENCE OF 1319-3565 FROM N.A. (ISOFORM 3).  
 CC TISSUE=Brain;  
 CC RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,  
 CC Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,  
 CC Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
 CC Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 CC Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,  
 CC Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 CC "NEDO human cDNA sequencing project.";   
 CC Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
 CC (3)  
 CC SEQUENCE OF 825-3565 FROM N.A. (ISOFORM 4).  
 CC MEDLINE=22897296; PubMed=12975309;  
 CC Clark H.F., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,  
 CC Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,  
 CC Eaton D., Foster J., Grimaldi C., Gu Q., Hass P.E., Heldens S.,  
 CC Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,  
 CC Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,  
 CC Seshagiri S., Simmons L., Singh J., Smith J., Stinson J., Vagts A.,  
 CC Vandlen R., Watanabe C., Wiedand D., Woods K., Xie M.-H., Yansura D.,  
 CC Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,  
 CC Godowski P.;  
 CC "The secreted protein discovery initiative (SPDI), a large-scale  
 CC effort to identify novel human secreted and transmembrane proteins: a  
 CC bioinformatics assessment.";   
 CC Genome Res. 13:2265-2270(2003).  
 CC (4)  
 CC SEQUENCE OF 966-2013 FROM N.A. (ISOFORM 4), AND TISSUE SPECIFICITY.  
 CC TISSUE=Brain;  
 CC MEDLINE=21456161; PubMed=11572484;  
 CC Nagase T., Kikuno R., Ohara O.;  
 CC "Prediction of the coding sequences of unidentified human genes. XXI.  
 CC The complete sequences of 60 new cDNA clones from brain which code for  
 CC large proteins.";   
 CC DNA Res. 8:179-187(2001).  
 CC (5)  
 CC DISEASE.  
 CC MEDLINE=22581359; PubMed=12696061;  
 CC Toomes C., Jackson A., Maguire K., Wood J., Gollin S., Ishwad C.,  
 CC Paterson I., Prime S., Parkinson K., Bell S., Woods G., Markham A.,  
 CC Oliver R., Woodward R., Sloan P., Dixon M., Read A., Thakker N.;  
 CC "The presence of multiple regions of homozygous deletion at the CSMD1  
 CC locus in oral squamous cell carcinoma question the role of CSMD1 in  
 CC head and neck carcinogenesis.";   
 CC Genes Chromosomes Cancer 37:132-140(2003).  
 CC (6)  
 CC DISEASE.  
 CC MEDLINE=22868729; PubMed=14506705;  
 CC Scholnick S.B., Richter T.M.;  
 CC "The role of CSMD1 in head and neck carcinogenesis.";   
 CC Genes Chromosomes Cancer 38:281-283(2003).  
 CC -!- FUNCTION: Potential suppressor of squamous cell carcinomas.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=Q96PZ7-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Short;  
 CC IsoId=Q96PZ7-2; Sequence=VSP\_009034, VSP\_009035;  
 CC Name=3;  
 CC IsoId=Q96PZ7-3; Sequence=VSP\_009030, VSP\_009031;  
 CC Note=No experimental confirmation available;

CC Name=4;  
 CC IsoId=Q96PZ7-4; Sequence=VSP\_009032, VSP\_009033;  
 CC -!- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in  
 CC brain. Expressed at intermediate level in brain, including  
 CC cerebellum, substantia nigra, hippocampus and fetal brain.  
 CC -!- DISEASE: Defects in CSMD1 may be a cause of oral and oropharyngeal  
 CC squamous cell carcinomas (OSCCs). Ref.5 and Ref.6 are however in  
 CC disagreement: while Ref.6 considers CSMD1 as a strong candidate  
 CC for OSCCs, Ref.5 thinks it is not.  
 CC -!- SIMILARITY: Belongs to the CSMD family.  
 CC -!- SIMILARITY: Contains 14 CUB domains  
 CC -!- SIMILARITY: Contains 28 Sushi (SCR) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF333704; AAK73475.2; -;  
 CC EMBL; AY017307; AAG52948.1; -;  
 CC EMBL; AX126936; BAC86754.1; ALT\_INIT.  
 CC EMBL; AY358174; AAG88541.1; ALT\_INIT.  
 CC EMBL; AB067477; BAB67783.1; -;  
 CC Genew; HGNC:14026; CSMD1.  
 CC InterPro; IPR000859; CUB.  
 CC InterPro; IPR000436; Sushi\_SCR\_CCP.  
 CC Pfam; PF00431; CUB; 14.  
 CC Pfam; PF00084; sushi; 27.  
 CC SMART; SM00032; CCF; 24.  
 CC SMART; SM00042; CUB; 14.  
 CC PROSITE; PS01180; CUB; 14.  
 CC Repeat; Signal; Transmembrane; Sushi; Alternative splicing.  
 CC SIGNAL 1 26  
 CC CHAIN 27 3565  
 CC POTENTIAL.  
 CC CUB AND SUSHI MULTIPLE DOMAINS PROTEIN 1.  
 CC FT DOMAIN 27 3488  
 CC EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 3489 3509  
 CC POTENTIAL.  
 CC FT DOMAIN 3510 3565  
 CC CYTOPLASMIC (POTENTIAL).  
 CC FT CUB 1 140  
 CC CUB 1.  
 CC FT SUSHI 1 145  
 CC SUSHI 1.  
 CC FT CUB 2 208  
 CC CUB 2.  
 CC FT SUSHI 2 349  
 CC SUSHI 2.  
 CC FT CUB 3 412  
 CC CUB 3.  
 CC FT SUSHI 3 528  
 CC SUSHI 3.  
 CC FT CUB 4 585  
 CC CUB 4.  
 CC FT SUSHI 4 698  
 CC SUSHI 4.  
 CC FT CUB 5 759  
 CC CUB 5.  
 CC FT SUSHI 5 874  
 CC SUSHI 5.  
 CC FT CUB 6 931  
 CC CUB 6.  
 CC FT SUSHI 6 1046  
 CC SUSHI 6.  
 CC FT CUB 7 1105  
 CC CUB 7.  
 CC FT SUSHI 7 1218  
 CC SUSHI 7.  
 CC FT CUB 8 1278  
 CC CUB 8.  
 CC FT SUSHI 8 1392  
 CC SUSHI 8.  
 CC FT CUB 9 1452  
 CC CUB 9.  
 CC FT SUSHI 9 1565  
 CC SUSHI 9.  
 CC FT CUB 10 1626  
 CC CUB 10.  
 CC FT SUSHI 10 1742  
 CC SUSHI 10.  
 CC FT CUB 11 1803  
 CC CUB 11.  
 CC FT SUSHI 11 1916  
 CC SUSHI 11.  
 CC FT CUB 12 2088  
 CC CUB 12.  
 CC FT SUSHI 12 2143  
 CC SUSHI 12.  
 CC FT CUB 13 2147  
 CC CUB 13.  
 CC FT SUSHI 13 2259  
 CC SUSHI 13.  
 CC FT CUB 14 2320  
 CC CUB 14.  
 CC FT SUSHI 14 2433  
 CC SUSHI 14.  
 CC FT CUB 15 2496  
 CC CUB 15.  
 CC FT SUSHI 15 2558  
 CC SUSHI 15.  
 CC FT CUB 16 2623  
 CC CUB 16.  
 CC FT SUSHI 16 2676  
 CC SUSHI 16.  
 CC FT CUB 17 2681  
 CC CUB 17.  
 CC FT SUSHI 17 2734  
 CC SUSHI 17.  
 CC FT CUB 18 2739  
 CC CUB 18.  
 CC FT SUSHI 18 2797  
 CC SUSHI 18.  
 CC FT CUB 19 2797  
 CC CUB 19.  
 CC FT SUSHI 19 2797  
 CC SUSHI 19.



RA SEQUENCE OF 77-1627 FROM N.A., SEQUENCE OF 81-98; 117-126; 210-224;  
 RP 466-485; 507-519; 576-593; 609-621; 718-736; 742-754; 1006-1017;  
 RP 1259-1273; 1369-1374; 1389-1398; 1490-1509; 1524-1533 AND 1537-1544,  
 RP VARIANT SR-944, AND TISSUE SPECIFICITY.  
 RC TISSUE-Placenta, and Serum;  
 RX MEDLINE=94146014; PubMed=7508748;  
 RA Kristensen T., Oxvig C., Sand O., Moller N.P.H., Sottrup-Jensen L.;  
 RT "Amino acid sequence of human pregnancy-associated plasma protein-A  
 RT derived from cloned cDNA";  
 RL Biochemistry 33:1592-1598(1994).  
 RN [3]  
 RP SEQUENCE OF 81-89; 117-126; 210-224; 460-485; 507-519; 576-593;  
 RP 718-736; 742-754; 1259-1273; 1369-1374; 1490-1509; 1524-1533 AND  
 RP 1537-1544, SUBUNITS, AND INTERCHAIN DISULFIDE BOND.  
 RC TISSUE=Serum;  
 RX MEDLINE=93286045; PubMed=7685339;  
 RA Oxvig C., Sand O., Kristensen T., Gleich G.J., Sottrup-Jensen L.;  
 RT "Circulating human pregnancy-associated plasma protein-A is disulfide-  
 RT bridged to the proform of eosinophil major basic protein.";  
 RL J. Biol. Chem. 268:12243-12246(1993).  
 RN [4]  
 RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.  
 RX MEDLINE=22421368; PubMed=12421832;  
 RA Overgaard M.T., Sorensen E.S., Stachowiak D., Boldt H.B.,  
 RA Kristensen L., Sottrup-Jensen L., Oxvig C.;  
 RT "Complex of pregnancy-associated plasma protein-A and the proform of  
 RT eosinophil major basic protein. Disulfide structure and carbohydrate  
 RT attachment sites.";  
 RL J. Biol. Chem. 278:2106-2117(2003).  
 RN [5]  
 RP IDENTIFICATION, FUNCTION, SUBCELLULAR LOCATION, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=99179030; PubMed=10077652;  
 RA Lawrence J.B., Oxvig C., Overgaard M.T., Sottrup-Jensen L.,  
 RA Gleich G.J., Hays L.G., Yates J.R. III, Conover C.A.;  
 RT "The insulin-like growth factor (IGF)-dependent IGF binding protein-4  
 RT protease secreted by human fibroblasts is pregnancy-associated plasma  
 RT protein-A";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3149-3153(1999).  
 RN [6]  
 RP FUNCTION, SUBUNITS, AND ENZYME REGULATION.  
 RX MEDLINE=20469470; PubMed=10913121;  
 RA Overgaard M.T., Haaning J., Boldt H.B., Olsen I.M., Laursen L.S.,  
 RA Christiansen M., Gleich G.J., Sottrup-Jensen L., Conover C.A.,  
 RA Oxvig C.;  
 RT "Expression of recombinant human pregnancy-associated plasma protein-A  
 RT and identification of the proform of eosinophil major basic protein  
 RT as its physiological inhibitor";  
 RL J. Biol. Chem. 275:31128-31133(2000).  
 RN [7]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=95057018; PubMed=7526035;  
 RA Bonno M., Oxvig C., Kephart G.M., Wagner J.M., Kristensen T.,  
 RA Sottrup-Jensen L., Gleich G.J.;  
 RT "Localization of pregnancy-associated plasma protein-A and  
 RT colocalization of pregnancy-associated plasma protein-A and  
 RT ribonucleic acid and eosinophil granule major basic protein messenger  
 RT ribonucleic acid in placenta";  
 RL Lab. Invest. 71:560-566(1994).  
 RN [8]  
 RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RX MEDLINE=99423540; PubMed=10491647;  
 RA Overgaard M.T., Oxvig C., Christiansen M., Lawrence J.B.,  
 RA Conover C.A., Gleich G.J., Sottrup-Jensen L., Haaning J.;  
 RT "Messenger ribonucleic acid levels of pregnancy-associated plasma  
 RT protein-A and the proform of eosinophil major basic protein:  
 RT expression in human reproductive and nonreproductive tissues.";  
 RL Biol. Reprod. 61:1083-1089(1999).  
 RN [9]  
 RP DEVELOPMENTAL STAGE.  
 RX MEDLINE=95293954; PubMed=7539791;

RA Oxvig C., Haaning J., Kristensen L., Wagner J.M., Rubin I.,  
 RA Stigbrand T., Gleich G.J., Sottrup-Jensen L.;  
 RT "Identification of angiotensinogen and complement C3dg as novel  
 RT proteins binding the proform of eosinophil major basic protein in  
 RT human pregnancy serum and plasma";  
 RL J. Biol. Chem. 270:13645-13651(1995).  
 CC -!- FUNCTION: Metalloproteinase which specifically cleaves IGFBP-4 in  
 CC the presence of IGF, resulting in release of bound IGF.  
 CC -!- CATALYTIC ACTIVITY: Cleavage of the 135-Met-Lys-136 bond in  
 CC insulin-like growth factor binding protein (IGFBP)-4, and the 143-  
 CC Ser-Lys-144 bond in IGFBP-5.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- ENZYME REGULATION: Inhibited by complexation with the proform  
 CC of PRG2.  
 CC -!- SUBUNIT: Homodimer; disulfide-linked. In pregnancy serum,  
 CC predominantly found as a disulfide-linked 2:2 heterotetramer with  
 CC the proform of PRG2.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: High levels in placenta and pregnancy serum.  
 CC In placenta, expressed in X cells in septa and anchoring villi,  
 CC and in syncytiotrophoblasts in the chorionic villi. Lower levels  
 CC are found in a variety of other tissues including kidney,  
 CC myometrium, endometrium, ovaries, breast, prostate, bone marrow,  
 CC colon, fibroblasts and osteoblasts.  
 CC -!- DEVELOPMENTAL STAGE: Present in serum and placenta during  
 CC pregnancy; levels increase throughout pregnancy.  
 CC -!- INDUCTION: By 8-bromoadenosine-3',5'-phosphate.  
 CC -!- PM: There appear to be no free sulfhydryl groups.  
 CC -!- SIMILARITY: Contains 5 Sushi (SCR) domains.  
 CC -!- SIMILARITY: Belongs to peptidase family M43B.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U28727; AAC50543.1; -;  
 DR EMBL; X68280; CAA48341.1; -;  
 DR PIR; S65464; S65464.  
 DR MEROPS; M43.004; -;  
 DR Genew; HGNC:8602; PAPPa.  
 DR MIM; 176385; -;  
 DR GO; GO:0005576; C:extracellular; IDA.  
 DR GO; GO:0008237; F:metallopeptidase activity; IDA.  
 DR GO; GO:0008270; F:zinc ion binding; NAS.  
 DR GO; GO:0007565; P:pregnancy; NAS.  
 DR InterPro; IPR008985; ConA\_like\_lect\_gl.  
 DR InterPro; IPR006558; LamG like.  
 DR InterPro; IPR000800; Notch\_dom.  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR InterPro; IPR008754; Peptidase\_M43B.  
 DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF05572; Peptidase\_M46; 1.  
 DR Pfam; PF00084; sushi; 4.  
 DR SMART; SM00032; CCP; 4.  
 DR SMART; SM00560; LamGL; 1.  
 DR SMART; SM00004; NL; 3.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Metal-binding; Zinc; Signal; Glycoprotein;  
 KW Zymogen; Repeat; Sushi.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT PROPEP 23 80  
 FT CHAIN 81 1627 PAPPALYSIN-1.  
 FT DOMAIN 24 83 ARG-RICH.  
 FT DOMAIN 272 583 PROTEOLYTIC.  
 FT DOMAIN 1215 1280 SUSHI 1.  
 FT DOMAIN 1285 1342 SUSHI 2.  
 FT DOMAIN 1346 1410 SUSHI 3.  
 FT DOMAIN 1415 1471 SUSHI 4.  
 FT DOMAIN 1478 1554 SUSHI 5.

FT	METAL	562	562	ZINC (CATALYTIC) (BY SIMILARITY).	RL	Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
FT	ACT SITE	563	563	BY SIMILARITY.	RN	[3]
FT	METAL	566	566	ZINC (CATALYTIC) (BY SIMILARITY).	RP	SEQUENCE FROM N.A.
FT	DISULFID	144	235		RA	Wallis J., Brown A.;
FT	DISULFID	327	622		RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
FT	DISULFID	332	657		RN	[4]
FT	DISULFID	414	428		RP	SEQUENCE FROM N.A. (ISOFORM 3).
FT	DISULFID	424	440		RC	TISSUE=Brain;
FT	DISULFID	457	473		RX	MEDLINE=22388257; PubMed=12477932;
FT	DISULFID	461	461		RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
FT	DISULFID	474	485	INTERCHAIN (WITH C-51 IN PRG2 PROFORM).	RA	Klausner R.D., Collins F.S., Wagner L., Stenmen C.M., Schuler G.D.,
FT	DISULFID	583	610	OR 583-612.	RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
FT	DISULFID	587	602	OR 587-600.	RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
FT	DISULFID	710	878		RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
FT	DISULFID	713	881		RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
FT	DISULFID	732	732		RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
FT	DISULFID	753	835		RA	Raha S.S., Igoellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
FT	DISULFID	775	781		RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
FT	DISULFID	947	975		RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
FT	DISULFID	960	971		RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
FT	DISULFID	983	990		RA	Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
FT	DISULFID	999	1011		RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
FT	DISULFID	1036	1070		RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
FT	DISULFID	1051	1139		RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
FT	DISULFID	1192	1205		RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
FT	DISULFID	1210	1210	INTERCHAIN.	RA	Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
FT	DISULFID	1215	1269		RT	"Generation and initial analysis of more than 15,000 full-length
FT	DISULFID	1227	1238		RL	human and mouse cDNA sequences.";
FT	DISULFID	1242	1280		RN	[5]
FT	DISULFID	1285	1329		RP	SEQUENCE OF 2542-3487 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
FT	DISULFID	1300	1310		RC	TISSUE=Brain;
QY	1 CVP LTC 6				RX	MEDLINE=21456161; PubMed=11572484;
Db	1410 CVP VTC 1415				RA	Nagase T., Kikuno R., Ohara O.;
					RT	"Prediction of the coding sequences of unidentified human genes. XXI.
					RT	The complete sequences of 60 new cDNA clones from brain which code for
					RT	large proteins.";
					RL	DNA Res. 8:179-187(2001).
					CC	-!- ALTERNATIVE PRODUCTS:
					CC	Event-Alternative splicing; Named isoforms=3;
					CC	Name=1;
					CC	Isoid=Q7Z408-1; Sequence=Displayed;
					CC	Name=2;
					CC	Isoid=Q7Z408-2; Sequence=VSP_009038, VSP_009039, VSP_009040,
					CC	VSP_009041, VSP_009042;
					CC	Notes=No experimental confirmation available;
					CC	Name=3;
					CC	Isoid=Q7Z408-3; Sequence=VSP_009043, VSP_009044, VSP_009045,
					CC	VSP_009046;
					CC	Note=No experimental confirmation available;
					CC	-!- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in
					CC	brain. Expressed at intermediate level in brain, including
					CC	cerebellum, substantia nigra, hippocampus and fetal brain.
					CC	Overexpressed in some head and neck cancer cell lines.
					CC	-!- SIMILARITY: Belongs to the CSMD family.
					CC	-!- SIMILARITY: Contains 14 CUB domains.
					CC	-!- SIMILARITY: Contains 26 Sushi (SCR) domains.
					CC	-!- CAUTION: Ref.3 sequence differs from that shown due to a
					CC	frameshift in position 939 that shortens the protein by 1021
					CC	residues in its N-terminus. It is unknown whether the sequence
					CC	shown exists or whether Ref.3 is right, shortening the sequence in
					CC	its N-terminus.
					CC	-!- CAUTION: Ref.3 sequence differs from that shown due to erroneous
					CC	gene model prediction.
					CC	-----
					CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
					CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
					CC	European Bioinformatics Institute. There are no restrictions on its
					CC	use by non-profit institutions as long as its content is in no way
					CC	modified and this statement is not removed. Usage by and for commercial
					CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
					CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
					CC	-----









FT SIGNAL 1 19  
 FT CHAIN 20 627  
 FT PROPEP 628 650  
 FT  
 FT DOMAIN 20 66  
 FT DOMAIN 67 104  
 FT DOMAIN 205 247  
 FT DOMAIN 251 292  
 FT DOMAIN 291 335  
 FT DOMAIN 482 530  
 FT DOMAIN 531 568  
 FT DOMAIN 601 627  
 FT DISULFID 24 37  
 FT DISULFID 32 49  
 FT DISULFID 51 65  
 FT DISULFID 71 81  
 FT DISULFID 76 91  
 FT DISULFID 93 103  
 FT DISULFID 209 222  
 FT DISULFID 218 231  
 FT DISULFID 233 246  
 FT DISULFID 255 269  
 FT DISULFID 263 278  
 FT DISULFID 280 291  
 FT DISULFID 295 307  
 FT DISULFID 300 316  
 FT DISULFID 318 334  
 FT DISULFID 486 500  
 FT DISULFID 492 516  
 FT DISULFID 518 529  
 FT DISULFID 535 550  
 FT DISULFID 543 559  
 FT DISULFID 561 567  
 FT LPID 627 627  
 FT  
 FT CARBOHYD 141 141  
 FT CARBOHYD 182 182  
 FT CARBOHYD 348 348  
 FT CARBOHYD 382 382  
 FT VARIANT 235 237  
 FT VARIANT 507 507  
 SQ SEQUENCE 650 AA; 71721 MW; 5DED71E354D8312A CRC64;  
 Query Match 86.8%; Score 33; DB 1; Length 650;  
 Best Local Similarity 83.3%; Pred. No. 30;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVPLTC 6  
 Db 147 CVPTTC 152  
 RESULT 9  
 LMG3 HUMAN STANDARD; PRT; 1587 AA.  
 AC Q9Y6N6;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Laminin gamma-3 chain precursor (Laminin 12 gamma 3).  
 GN LAMC3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99242614; PubMed=10225960;  
 RA Koch M., Olson P.F., Albus A., Jin W., Hunter D.D., Brunken W.J.,  
 RA Burgeson R.E., Champilaud M.F.;  
 RT "Characterization and expression of the laminin gamma3 chain: a novel,  
 RT non-basement membrane-associated, laminin chain.";

RL J. Cell Biol. 145:605-618(1999).  
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin  
 CC is thought to mediate the attachment, migration and organization  
 CC of cells into tissues during embryonic development by interacting  
 CC with other extracellular matrix components.  
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
 CC different polypeptide chains (alpha, beta, gamma), which are bound  
 CC to each other by disulfide bonds into a cross-shaped molecule  
 CC comprising one long and three short arms with globules at each  
 CC end. The gamma-3 chain is a subunit of laminin-12.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- TISSUE SPECIFICITY: Broadly expressed in: skin, heart, lung, and  
 CC the reproductive tracts.  
 CC -!- DOMAIN: The alpha-helical domains I and II are thought to interact  
 CC with other laminin chains to form a coiled coil structure.  
 CC -!- DOMAIN: Domain IV is globular.  
 CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.  
 CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.  
 CC -!- SIMILARITY: Contains 1 laminin IV domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC -----  
 CC EMBL; AF041835; AAD36991.1; .  
 CC HSSP; P02468; 1TLE.  
 CC Genew; HGNC:6494; LAMC3.  
 CC MIM; 604349; .  
 CC GO; GO:0005578; C:extracellular matrix; TAS.  
 CC GO; GO:0016020; C:membrane; TAS.  
 CC GO; GO:0005198; F:structural molecule activity; TAS.  
 CC InterPro; IPR006209; EGF-like.  
 CC InterPro; IPR008212; Lam\_N2.  
 CC InterPro; IPR000034; Laminin B.  
 CC InterPro; IPR002049; Laminin\_EGF.  
 CC InterPro; IPR008211; LamNT.  
 CC Pfam; PF00052; laminin B; 1.  
 CC Pfam; PF00053; laminin EGF; 9.  
 CC Pfam; PF00055; laminin\_Nterm; 1.  
 CC PRINTS; PR00011; EGF\_LAMININ.  
 CC ProDom; PD002082; Lam\_N2; 1.  
 CC SMART; SM00180; EGF\_Lam; 9.  
 CC SMART; SM00136; LamNT; 1.  
 CC PROSITE; PS00022; EGF\_1; 7.  
 CC PROSITE; PS01186; EGF\_2; 2.  
 CC PROSITE; PS01248; LAMININ TYPE EGF; 10.  
 CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1587  
 FT DOMAIN 20 270  
 FT DOMAIN 271 326  
 FT DOMAIN 327 382  
 FT DOMAIN 383 429  
 FT DOMAIN 430 479  
 FT DOMAIN 480 489  
 FT DOMAIN 490 672  
 FT DOMAIN 673 706  
 FT DOMAIN 707 754  
 FT DOMAIN 755 809  
 FT DOMAIN 810 865  
 FT DOMAIN 866 916  
 FT DOMAIN 917 964  
 FT DOMAIN 965 1013  
 FT DOMAIN 1014 1587  
 FT DOMAIN 1071 1141  
 FT DOMAIN 1200 1229  
 FT DOMAIN 1424 1504  
 FT DOMAIN 1535 1579  
 FT  
 FT LAMININ GAMMA-3 CHAIN.  
 FT LAMININ N-TERMINAL (DOMAIN VI).  
 FT LAMININ EGF-LIKE 1.  
 FT LAMININ EGF-LIKE 2.  
 FT LAMININ EGF-LIKE 3.  
 FT LAMININ EGF-LIKE 4.  
 FT LAMININ EGF-LIKE 5 (N-TERMINAL).  
 FT LAMININ DOMAIN IV.  
 FT LAMININ EGF-LIKE 5 (C-TERMINAL).  
 FT LAMININ EGF-LIKE 6.  
 FT LAMININ EGF-LIKE 7.  
 FT LAMININ EGF-LIKE 8.  
 FT LAMININ EGF-LIKE 9.  
 FT LAMININ EGF-LIKE 10.  
 FT LAMININ EGF-LIKE 11.  
 FT DOMAIN II AND I.  
 FT COILED COIL (POTENTIAL).  
 FT COILED COIL (POTENTIAL).  
 FT COILED COIL (POTENTIAL).  
 FT COILED COIL (POTENTIAL).

FT SITE 1059 1061 CELL ATTACHMENT SITE (POTENTIAL).  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 837 837 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1185 1185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1518 1518 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1587 AA; 172051 MW; 3CB6E09B5F203319 CRC64;  
Query Match 86.8%; Score 33; DB 1; Length 1587;  
Best Local Similarity 83.3%; Pred. No. 73; Indels 0; Gaps 0;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTC 6  
Db 704 CVPTC 709  
RESULT 10  
VANB\_PSES9  
ID VANB\_PSES9 STANDARD; PRT; 314 AA.  
AC P12530;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Vanillate O-demethylase oxidoreductase (EC 1.14.13.-) (Vanillate degradation ferredoxin-like protein).  
GN VANB  
OS Pseudomonas sp. (strain ATCC 19151).  
OC Bacteria; Proteobacteria.  
OX NCBI\_TaxID=315;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89008117; PubMed=3170489;  
RA Brunel F., Davison J.;  
RT "Cloning and sequencing of Pseudomonas genes encoding vanillate demethylase."  
RL J. Bacteriol. 170:4924-4930(1988).  
CC -!- COPACTOR: FMN (By similarity).  
CC -!- PATHWAY: Vanillate degradation (vanillate is a key intermediate in the degradation of lignin).  
CC -!- SIMILARITY: Belongs to the PDR/vanB family.  
CC -!- SIMILARITY: In the C-terminal section; belongs to the 2Fe2S plant-type ferredoxin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M22077; AAA26020.1; .  
DR HSSP; F33164; 2PIA.  
DR InterPro; IPR006058; 2Fe2S\_fd\_BS.  
DR InterPro; IPR008333; FAD binding\_6.  
DR InterPro; IPR001041; Ferredoxin.  
DR InterPro; IPR001433; Oxred\_FAD/NAD(P).  
DR InterPro; IPR000951; PhioX\_reductase.  
DR Pfam; PF00970; FAD\_binding\_5; 1.  
DR Pfam; PF00111; fer2; 1.  
DR Pfam; PF00175; NAD binding\_1; 1.  
DR PRINTS; PR00409; PHDIOXRDTASE.  
DR PROSITE; PS00197; 2Fe2S\_FERREDOXIN; 1.  
KW Aromatic hydrocarbons catabolism; Lignin degradation; Oxidoreductase;  
KW Flavoprotein; FMN; NAD; Metal-binding; Iron-sulfur; Iron; 2Fe-2S;  
KW Electron transport.  
FT NP\_BIND 1 95 FMN.  
FT NP\_BIND 105 217 NAD.

FT METAL 263 263 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 268 268 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 271 271 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
FT METAL 302 302 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
SQ SEQUENCE 314 AA; 33706 MW; FC521516AA6CEB72 CRC64;  
Query Match 84.2%; Score 32; DB 1; Length 314;  
Best Local Similarity 66.7%; Pred. No. 23; Indels 0; Gaps 0;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTC 6  
Db 258 CIPLAC 263  
RESULT 11  
APOH\_BOVIN  
ID APOH\_BOVIN STANDARD; PRT; 345 AA.  
AC P17690; Q28052;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI) (Beta(2)GPI).  
GN APOH.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Gao B., Virmani M., Romm E., Lazar-Wesley E., Sakaguchi K., Appella E., Kunos G., Takacs L.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 4-345 FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
RC TISSUE=Liver;  
RX MEDLINE=92232647; PubMed=1567819;  
RA Bendixen E., Halkier T., Magnusson S., Sottrup-Jensen L., Kristensen T.;  
RT "Complete primary structure of bovine beta 2-glycoprotein I: localization of the disulfide bridges."  
RL Biochemistry 31:3611-3617(1992).  
RN [3]  
RP SEQUENCE OF 20-345, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.  
RC TISSUE=Plasma;  
RX MEDLINE=92089075; PubMed=1751487;  
RA Kato H., Enjyoji K.-I.;  
RT "Amino acid sequence and location of the disulfide bonds in bovine beta 2 glycoprotein I: the presence of five Sushi domains."  
RL Biochemistry 30:11687-11694(1991).  
RN [4]  
RP SEQUENCE OF 20-41.  
RX MEDLINE=90226328; PubMed=2327984;  
RA Li Q., Blacher R., Esch F., Congote L.F.;  
RT "Isolation from fetal bovine serum of an apolipoprotein-H-like protein which inhibits thymidine incorporation in fetal calf erythroid cells."  
RL Biochem. J. 267:261-264 (1990).  
CC -!- FUNCTION: Binds to various kinds of negatively charged substances such as heparin, phospholipids, and dextran sulfate. May prevent activation of the intrinsic blood coagulation cascade by binding to phospholipids on the surface of damaged cells.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.  
CC -!- SIMILARITY: Contains 4 Sushi (SCR) domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC      EMBL; L07303; AAA30382.1; -.
CC      EMBL; X60065; CAA42669.1; -.
CC      PIR; JN0502; NBBO.
CC      HSSP; P02749; 1C1Z.
CC      InterPro; IPR000436; Sushi_SCR_CCP.
CC      Pfam; PF00084; sushi; 4.
CC      SMART; SM00032; CCP; 4.
CC      Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal.
FT      SIGNAL 1 19
FT      CHAIN 20 345 BETA-2-GLYCOPROTEIN I.
FT      DOMAIN 22 80 SUSHI 1.
FT      DOMAIN 83 138 SUSHI 2.
FT      DOMAIN 141 201 SUSHI 3.
FT      DOMAIN 204 261 SUSHI 4.
FT      DOMAIN 263 345 SUSHI-LIKE.
FT      DISULFID 23 66
FT      DISULFID 51 79
FT      DISULFID 84 124
FT      DISULFID 110 137
FT      DISULFID 142 188
FT      DISULFID 174 200
FT      DISULFID 205 248
FT      DISULFID 234 260
FT      DISULFID 300 325
FT      DISULFID 307 345
FT      CARBOHYD 92 92 N-LINKED (GLCNAC. . .)
FT      CARBOHYD 162 162 N-LINKED (GLCNAC. . .)
FT      CARBOHYD 183 183 N-LINKED (GLCNAC. . .)
FT      CARBOHYD 193 193 N-LINKED (GLCNAC. . .)
FT      CARBOHYD 253 253 N-LINKED (GLCNAC. . .)
FT      CONFLICT 101 101 E -> G (IN REF. 1).
FT      CONFLICT 108 108 F -> S (IN REF. 1).
FT      CONFLICT 177 177 H -> R (IN REF. 1).
FT      CONFLICT 191 191 H -> N (IN REF. 3).
FT      CONFLICT 194 194 W -> C (IN REF. 1).
FT      CONFLICT 259 259 S -> N (IN REF. 1).
FT      CONFLICT 302 302 H -> N (IN REF. 1).
FT      CONFLICT 305 305 K -> R (IN REF. 1).
FT      CONFLICT 329 329 H -> R (IN REF. 1).
SQ      SEQUENCE 345 AA; 38252 MW; E117DAB609461G33 CRC64;

Query Match 84.2%; Score 32; DB 1; Length 345;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVPLTC 6
Db      137 CAPITC 142

RESULT 12
APOH PANTR STANDARD; PRT; 345 AA.
AC Q95LE0;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Beta-2-glycoprotein I precursor (Apolipoprotein H) (Apo-H) (B2GPI)
DE (Beta(2)GPI).
GN APOH
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT GLU-229.
RX MEDLINE=21372074; PubMed=11479737;
RA Sanghera D.K., Nestlode C.S., Ferrell R.E., Kamboh M.I.;

"Chimpanzee apolipoprotein H (beta2-glycoprotein I): report on the gene structure, a common polymorphism, and a high prevalence of antiphospholipid antibodies.";
Hum. Genet. 109:63-72(2001).
-!- FUNCTION: Binds to various kinds of negatively charged substances such as heparin, phospholipids, and dextran sulfate. May prevent activation of the intrinsic blood coagulation cascade by binding to phospholipids on the surface of damaged cells.
-!- SUBCELLULAR LOCATION: Secreted (By similarity).
-!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
-!- SIMILARITY: Contains 4 Sushi (SCR) domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```

RT      EMBL; AF358415; AAK71538.1; -.
RT      EMBL; AF358408; AAK71538.1; JOINED.
RT      EMBL; AF358409; AAK71538.1; JOINED.
RT      EMBL; AF358410; AAK71538.1; JOINED.
RT      EMBL; AF358411; AAK71538.1; JOINED.
RT      EMBL; AF358412; AAK71538.1; JOINED.
RT      EMBL; AF358413; AAK71538.1; JOINED.
RT      EMBL; AF358414; AAK71538.1; JOINED.
RT      InterPro; IPR000436; Sushi_SCR_CCP.
RT      Pfam; PF00084; sushi; 4.
RT      SMART; SM00032; CCP; 4.
RT      Heparin-binding; Glycoprotein; Plasma; Repeat; Sushi; Signal;
RT      Polymorphism. 1 19
RT      CHAIN 20 345 BETA-2-GLYCOPROTEIN I.
RT      DOMAIN 22 80 SUSHI 1.
RT      DOMAIN 83 138 SUSHI 2.
RT      DOMAIN 141 201 SUSHI 3.
RT      DOMAIN 204 261 SUSHI 4.
RT      DOMAIN 263 345 SUSHI-LIKE.
RT      DISULFID 23 66 BY SIMILARITY.
RT      DISULFID 51 79 BY SIMILARITY.
RT      DISULFID 84 124 BY SIMILARITY.
RT      DISULFID 110 137 BY SIMILARITY.
RT      DISULFID 142 188 BY SIMILARITY.
RT      DISULFID 174 200 BY SIMILARITY.
RT      DISULFID 205 248 BY SIMILARITY.
RT      DISULFID 234 260 BY SIMILARITY.
RT      DISULFID 264 315 BY SIMILARITY.
RT      DISULFID 300 325 BY SIMILARITY.
RT      DISULFID 307 345 BY SIMILARITY.
RT      CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
RT      CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
RT      CARBOHYD 193 193 N-LINKED (GLCNAC. . .) (POTENTIAL).
RT      CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
RT      VARIANT 229 229 K -> E (in allele APOH*4).
SQ      SEQUENCE 345 AA; 38267 MW; 712A3EDA2AD4FD36 CRC64;

Query Match 84.2%; Score 32; DB 1; Length 345;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 CVPLTC 6
Db      137 CAPITC 142

RESULT 13
DCAM_TRYBB
ID DCAM_TRYBB STANDARD; PRT; 370 AA.
AC P50244;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE S-adenosylmethionine decarboxylase proenzyme (EC 4.1.1.50) (AdoMetDC)  
DE (samDC) [Contains: S-adenosylmethionine decarboxylase alpha chain; S-  
DE adenosylmethionine decarboxylase beta chain].  
OS Trypanosoma brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OC NCBI\_TaxID=5702;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=EATRO 164;  
RC Scott J.R., Ullman B.;  
RA "Molecular cloning and functional expression of the S-  
RT adenosylmethionine decarboxylase gene of Leishmania donovani and  
RT Trypanosoma brucei.";  
RT Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
RL -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine = (S-deoxy-5-  
CC adenosyl) (3-aminopropyl) methylsulfonium salt + CO(2).  
CC -1- COFACTOR: Pyruvoyl group.  
CC -1- PATHWAY: Decarboxylation of S-adenosylmethionine provides the  
CC aminopropyl moiety required for spermidine and spermine  
CC biosynthesis from putrescine.  
CC -1- SIMILARITY: Belongs to the eukaryotic AdoMetDC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U20092; AAA61969.1; -.  
DR HSP; P17707; LUEN.  
DR InterPro: IPR001985; SAM decarbox.  
DR Pfam: PF01536; SAM decarbox; 1.  
DR ProDom: PD02379; SAM decarbox; 1.  
DR TIGRFAMs: TIGR00535; SAM DCase; 1.  
DR PROSITE: PS01336; ADOMETDC; 1.  
KW Spermidine biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen.  
FT CHAIN 1 85  
FT S-ADENOSYLMETHIONINE DECARBOXYLASE BETA  
FT CHAIN (BY SIMILARITY).  
FT CHAIN (BY SIMILARITY).  
FT CLEAVAGE (NONHYDROLYTIC)  
FT (BY SIMILARITY).  
FT MOD\_RES 86 86  
FT CONVERTED TO A PYRUVYL GROUP  
FT (BY SIMILARITY).  
FT ACT\_SITE 29 29  
FT IMPORTANT FOR CATALYTIC ACTIVITY (BY  
FT SIMILARITY).  
FT ACT\_SITE 32 32  
FT IMPORTANT FOR CATALYTIC ACTIVITY (BY  
FT SIMILARITY).  
FT ACT\_SITE 100 100  
FT IMPORTANT FOR CATALYTIC ACTIVITY (BY  
FT SIMILARITY).  
SQ SEQUENCE 370 AA; 41692 MW; 06E0BF4BEAED4D42 CRC64;  
Query Match 84.2%; Score 32; DB 1; Length 370;  
Best Local Similarity 83.3%; Pred. No. 28;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTC 6  
DB 108 CVPLTC 113  
RESULT 14  
MUB1\_XENLA STANDARD; PRT; 398 AA.  
ID MUB1\_XENLA  
AC P38555;  
AC 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Integumentary mucin B.1 (FIM-B.1) (Fragment).  
OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OC NCBI\_TaxID=8355;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91002513; PubMed=2207068;  
RA Probst J.C., Gertzen E.-M., Hoffmann W.;  
RT "An integumentary mucin (FIM-B.1) from Xenopus laevis homologous with  
RT von Willebrand factor.";  
RL Biochemistry 29:6240-6244 (1990).  
CC -1- FUNCTION: Could be involved in defense against microbial  
CC infections. Protects the epithelia from external environment.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- TISSUE SPECIFICITY: Expressed and stored exclusively in mature  
CC mucous glands of the skin.  
CC -1- PFM: Extensively O-glycosylated.  
CC -1- SIMILARITY: Contains 1 VWFC domain.  
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; J02910; AAA49711.1; -.  
DR PIR; A35281; A35281.  
DR InterPro: IPR006208; Cys knot.  
DR InterPro: IPR006207; Cys\_knot\_C.  
DR InterPro: IPR001007; VWFC\_C.  
DR Pfam: PF00007; Cys\_knot; 1.  
DR SMART: SM00041; CT; 1.  
DR SMART: SM00214; VWC; 2.  
DR PROSITE: PS01185; CTCK\_1; 1.  
DR PROSITE: PS01225; CTCK\_2; 1.  
DR PROSITE: PS01208; VWFC\_1; 1.  
DR PROSITE: PS01208; VWFC\_2; 1.  
DR PROSITE: PS50184; VWFC\_2; 1.  
KW Repeat; Glycoprotein.  
FT NON\_TER 1 93  
FT DOMAIN 1 93  
FT REPEAT <1 7  
FT REPEAT 8 18  
FT REPEAT 19 29  
FT REPEAT 30 40  
FT REPEAT 41 51  
FT REPEAT 52 62  
FT REPEAT 66 76  
FT REPEAT 83 93  
FT REPEAT 169 234  
FT DOMAIN 301 392  
FT DOMAIN 301 356  
FT DISULFID 323 370  
FT DISULFID 332 386  
FT DISULFID 336 388  
FT DISULFID ? 391  
FT DISULFID ? 391  
SQ SEQUENCE 398 AA; 42101 MW; C06C86A805A3C3A4 CRC64;  
Query Match 84.2%; Score 32; DB 1; Length 398;  
Best Local Similarity 83.3%; Pred. No. 30;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTC 6  
DB 233 CEPLTC 238  
RESULT 15  
CHRD\_BRARE STANDARD; PRT; 940 AA.  
ID CHRD\_BRARE

AC 057472; Q9DEB8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Chordin precursor (Chordino protein).  
GN CHD.  
OS Brachydanio rerio (Zebrafish) (Danio rerio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
OC Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RN [2]  
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL  
RP STAGE.  
RC TISSUE=Gastrula;  
RX MEDLINE=98104254; PubMed=9441687;  
RA Miller-Bertoglio V.E., Fisher S., Sanchez A., Mullins M.C.,  
RA Halpern M.E.;  
RT "Differential regulation of chordin expression domains in mutant  
RT zebrafish.";  
RL Dev. Biol. 192:537-550(1997).  
RN [2]  
RN  
SEQUENCE OF 1-42 FROM N.A.  
FUJII R., Hibi M., Hirano T., Shimizu T.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Dorsalizing factor. Key developmental protein that  
CC dorsalizes early vertebrate embryonic tissues by binding to  
CC ventralizing TGF-beta family bone morphogenetic proteins (BMPs)  
CC and sequestering them in latent complexes (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
CC -!- TISSUE SPECIFICITY: During gastrulation, levels are highest in  
CC the organizer region. Also present in the developing brain and in  
CC paraxial mesoderm and ectoderm.  
CC -!- DEVELOPMENTAL STAGE: First detected shortly after the midblastula  
CC transition. Levels increase during gastrulation, persist through  
CC early somitogenesis, but then decrease and are gone by 24  
CC hours.  
CC -!- SIMILARITY: Belongs to the chordin family.  
CC -!- SIMILARITY: Contains 4 CHRD domains.  
CC -!- SIMILARITY: Contains 4 VWFC domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AF034606; AAB93485.1; -;  
DR EMBL; AB043968; BAB18642.1; -;  
DR ZFIN; ZDB-GENE-990415-33; chd.  
DR InterPro; IPR006559; SOG.  
DR InterPro; IPR001007; VWFC\_C.  
DR Pfam; PF00093; VWC; 4.  
DR SMART; SM00566; SOG; 3.  
DR SMART; SM00214; VWC; 4.  
DR PROSITE; PS00933; CHRD; 4.  
DR PROSITE; PS01208; VWFC 1; 3.  
DR PROSITE; PS0184; VWFC 2; 4.  
KW Developmental protein; Repeat; Glycoprotein; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 940 CHORDIN.  
FT DOMAIN 42 118 VWFC 1.  
FT DOMAIN 162 277 CHRD 1.  
FT DOMAIN 279 398 CHRD 2.  
FT DOMAIN 404 519 CHRD 3.  
FT DOMAIN 525 652 CHRD 4.  
FT DOMAIN 689 748 VWFC 2.  
FT DOMAIN 767 836 VWFC 3.  
FT DOMAIN 855 919 VWFC 4.  
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT

SQ SEQUENCE 940 AA; 104999 MW; B855CAF84F4623AC CRC64;  
Query Match 84.2%; Score 32; DB 1; Length 940;  
Best Local Similarity 83.3%; Pred. No. 69;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CVPLTC 6  
| ||||  
Db 815 CPPLTC 820

Search completed: September 5, 2004, 09:56:09  
Job time : 3.72727 secs

**this Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:46:09 ; Search time 13.5758 Seconds  
(without alignments)  
139.448 Million cell updates/sec

Title: US-09-761-636A-12  
Perfect score: 38  
Sequence: 1 CVPLTC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvirus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	97.4	447	5	Q963E7
2	37	97.4	447	5	O18280
3	37	97.4	849	15	Q7SKG4
4	36	94.7	533	4	Q9H4W4
5	36	94.7	3389	4	Q96QU9
6	36	94.7	3564	11	Q923L3
7	36	94.7	3566	4	Q96RM4
8	35	92.1	426	13	Q9DGM2
9	35	92.1	468	11	Q9UX57
10	35	92.1	578	11	Q80VW3
11	35	92.1	946	4	Q96Q03
12	35	92.1	1367	11	Q9ES06
13	35	92.1	1545	11	Q8R4K8
14	35	92.1	1574	11	Q8R4K7
15	35	92.1	3487	4	Q7Z408
16	34	89.5	152	10	Q8H347

17	34	89.5	245	10	Q8LXU4
18	34	89.5	367	5	Q9XU36
19	34	89.5	511	10	Q8S2A0
20	34	89.5	570	4	Q8N9T9
21	33	86.8	66	13	Q9PTM2
22	33	86.8	66	16	Q9A1P4
23	33	86.8	68	13	Q9PTM3
24	33	86.8	93	15	Q78161
25	33	86.8	168	13	Q90YB8
26	33	86.8	246	5	Q95YD9
27	33	86.8	265	16	Q8DKA0
28	33	86.8	277	12	Q9PVV4
29	33	86.8	395	10	Q9L1D1
30	33	86.8	415	13	Q91548
31	33	86.8	415	13	Q7ZXQ5
32	33	86.8	429	13	Q91650
33	33	86.8	443	5	Q8MQR1
34	33	86.8	467	16	Q8YN93
35	33	86.8	556	5	Q7YZB6
36	33	86.8	637	10	Q9S5Y5
37	33	86.8	637	10	Q49077
38	33	86.8	650	5	Q9Y0V1
39	33	86.8	664	5	Q8W387
40	33	86.8	852	15	Q8UL55
41	33	86.8	852	15	Q8UL54
42	33	86.8	875	5	Q8IM00
43	33	86.8	1007	13	Q90ZM3
44	33	86.8	1230	5	Q9VXH3
45	33	86.8	1230	5	Q24284

#### ALIGNMENTS

RESULT 1  
Q963E7 PRELIMINARY; PRT; 447 AA.  
AC Q963E7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative chemoreceptor GUR-3.  
GN GUR-3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Robertson H.M.;  
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF387606; AAK70489.1; -;  
DR GO; GO:0016020; C-membrane; IEA.  
DR GO; GO:0005549; Fodorant binding; IEA.  
DR GO; GO:0004984; Fodorant receptor activity; IEA.  
DR GO; GO:0004872; Fodorant receptor activity; IEA.  
DR GO; GO:0007608; Prolactin activity; IEA.  
DR InterPro; IPR004117; 7tm\_6.  
DR Pfam; PF02949; 7tm\_6; 1.  
KW Receptor.  
SQ SEQUENCE 447 AA; 51176 MW; 3AB7B4216F048BEB CRC64;

Query Match 97.4%; Score 37; DB 5; Length 447;  
Best Local Similarity 83.3%; Pred. No. 9.2;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
|:|:|:|  
Db 302 CIPLTC 307

RESULT 2

```

018280
ID O18280 PRELIMINARY; PRT; 447 AA.
AC O18280;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ZC504.5 protein.
GN ZC504.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditia.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z50029; CAA90343.3; -.
DR PIR; T27621; T27621.
DR WormPep; ZC504.5; CE30223.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005549; F:odorant binding; IEA.
DR GO; GO:0004984; F:olfactory receptor activity; IEA.
DR GO; GO:0007608; P:olfaction; IEA.
DR InterPro; IPR004117; 7tm_6.
DR Pfam; PF02949; 7tm_6; 1.
SQ SEQUENCE 447 AA; 51227 MW; E916BEAA7C14C80C CRC64;

Query Match 97.4%; Score 37; DB 5; Length 447;
Best Local Similarity 83.3%; Pred. No. 9.2;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 302 CIPLTC 307

RESULT 3
Q7SKG4 PRELIMINARY; PRT; 849 AA.
AC Q7SKG4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
[1]
RN SEQUENCE FROM N.A.
RA STRAIN=02TH.OUR1331;
RX PubMed=12908933;
RA Tovanaubtra S., Watanaveeradej V., Viputtikul K., De Souza M.,
RA Ritzak M.H., Suriyanon V., Jittiwutikarn J., Sriplienchan S.,
RA Nitsayaphan S., Benenson M.W., Sirisopana N., Renzullo P.O.,
RA Brown A.E., Robb M.L., Beyrer C., Celentano D.D., McNeil J.G.,
RA Birx D.L., Carr J.K., McCutchan F.E.;
RT "A New Circulating Recombinant Form, CRF15_01B, Reinforces the Linkage
between IDU and Heterosexual Epidemics in Thailand.";
RL AIDS Res. Hum. Retroviruses 19:561-567(2003).
DR EMBL; AF529572; AAQ09553.1; -.
KW Envelope protein.
SQ SEQUENCE 849 AA; 96536 MW; D7AB6F8B670BAED3 CRC64;

Query Match 97.4%; Score 37; DB 15; Length 849;
Best Local Similarity 83.3%; Pred. No. 16;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 302 CIPLTC 307

Query Match 94.7%; Score 36; DB 4; Length 533;
Best Local Similarity 83.3%; Pred. No. 17;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 353 CVPITC 358

RESULT 5
Q96QU9 PRELIMINARY; PRT; 3389 AA.
AC Q96QU9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CUB and sushi multiple domains protein 1 short form.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,
RA Sunwoo J.B., Gollin S.M., Scholnick S.B.;
RT "Transcript map of the 8p23 putative tumor suppressor region.";
RL Genomics 75:17-25(2001).
CC -!- SIMILARITY: CONTAINS 14 CUB DOMAINS.
DR EMBL; AV017307; AAG52948.1; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00431; CUB; 14.
DR Pfam; PF00084; sushi; 25.
DR SMART; SM00032; CCP; 25.
DR SMART; SM00042; CUB; 14.
DR PROSITE; PS01180; CUB; 14.

```





```
FT VARIANT 318 318 W -> *
SQ SEQUENCE 426 AA; 48770 MW; F92C0B0B3046E4A3 CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 13; Length 426;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
|:|||||
Db 265 CMPLTC 270

RESULT 9
Q9JK57 ID Q9JK57 PRELIMINARY; PRT; 468 AA.
AC Q9JK57;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pregnancy-associated plasma protein A (Fragment).
GN PAPPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RA Olesen C., Hansen C., Hayashizaki Y., Byskov A., Tommerup N.;
RT "Partial sequence of Mus musculus Mus musculus pregnancy-associated
RT plasma protein A (Pappa).";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF260433; AAF70319.1; -.
DR MEROPS; M46.001; -.
DR MGD; MGI:97479; Pappa.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch dom.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 1.
FT NON TER 1
SQ SEQUENCE 468 AA; 51134 MW; 85768E10D7D34FDA CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 11; Length 468;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
|:|||||
Db 251 CVPVTC 256

RESULT 10
Q80VM3 ID Q80VM3 PRELIMINARY; PRT; 578 AA.
AC Q80VM3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE BM423Pb.2 (Pregnancy-associated plasma protein A) (Fragment).
GN PAPPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Dunn M.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL691454; CAD62274.1; -.
DR GO; GO:0016020; C:membrane; IEA.
```

```
DR GO; GO:0030154; P:cell differentiation; IEA.
DR InterPro; IPR000800; Notch dom.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 1.
FT NON TER 1
SQ SEQUENCE 578 AA; 64140 MW; C9F6DA810133FED7 CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 11; Length 578;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
|:|||||
Db 361 CVPVTC 366

RESULT 11
Q96Q03 ID Q96Q03 PRELIMINARY; PRT; 946 AA.
AC Q96Q03;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA1884 (Fragment).
GN KIAA1884.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RL DNA Res. 8:179-187(2001).
DR EMBL; AB067471; BAB67777.1; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 12.
DR SMART; SM00032; CCP; 12.
DR PROSITE; PS00290; IG_MHC; 1.
DR Hypothetical protein.
KW NON TER 1
FT NON TER 1
SQ SEQUENCE 946 AA; 102543 MW; EDC903C6BD390B4A CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 4; Length 946;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
|:|||||
Db 47 CVPVTC 52

RESULT 12
Q9ES06 ID Q9ES06 PRELIMINARY; PRT; 1367 AA.
AC Q9ES06;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Hourvitz A., Kuwahara A., Hennebold J.D., Negishi H., Tanaka M.,
RA Widger A., Erickson G.F., Adashi E.V.;
RT "The regulated expression of insulin-like growth factor binding
RT protein-4-endopeptidase (pregnancy-associated plasma protein-A) in the
RT rodent ovary: A marker of the dominant follicle and of the corpus
RT luteum.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF258461; AAG09799.1; -.
DR MEROPS; M46_001; -.
DR MGD; MGI:97479; Pappa.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008000; Notch dom.
DR InterPro; IPR008754; Peptidase_M43B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF05572; Peptidase_M46; I.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 1367 AA; 152644 MW; A46BC5B0449042A5 CRC64;

Query Match 92.1%; Score 35; DB 11; Length 1367;
Best Local Similarity 83.3%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 1150 CVPVTC 1155

RESULT 13
Q8R4K8 PRELIMINARY; PRT; 1545 AA.
AC Q8R4K8
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pregnancy-associated plasma protein-A (Fragment).
GN PAPP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,
RA Haaning J., Sottrup-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-Ai)
RT with differential proteolytic activity.";
RL Eur. J. Biochem. 0:0-0(2002).
DR EMBL; AF439513; AAM12687.1; -.
DR MGD; MGI:97479; Pappa.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR008000; Notch dom.
DR InterPro; IPR008754; Peptidase_M43B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF05572; Peptidase_M46; I.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 1574 AA; 176105 MW; AE53F0187E6EFB7B CRC64;

Query Match 92.1%; Score 35; DB 11; Length 1574;
Best Local Similarity 83.3%; Pred. No. 71;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 1357 CVPVTC 1362

RESULT 15
Q7Z408 PRELIMINARY; PRT; 3487 AA.
ID Q7Z408
AC Q7Z408

```

```

DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 1545 AA; 172584 MW; B820BF9938245419 CRC64;

Query Match 92.1%; Score 35; DB 11; Length 1545;
Best Local Similarity 83.3%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 1328 CVPVTC 1333

RESULT 14
Q8R4K7 PRELIMINARY; PRT; 1574 AA.
AC Q8R4K7
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pregnancy-associated plasma protein-A variant (Fragment).
GN PAPP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Soe R., Overgaard M.T., Thomsen A.R., Laursen L.S., Olsen I.M.,
RA Haaning J., Sottrup-Jensen L., Haaning J., Giudice L.C., Conover C.A.,
RA Oxvig C.;
RT "Expression of recombinant murine PAPP-A and a novel variant (PAPP-Ai)
RT with differential proteolytic activity.";
RL Eur. J. Biochem. 0:0-0(2002).
DR EMBL; AF439514; AAM12688.1; -.
DR MGD; MGI:97479; Pappa.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0030154; P:cell differentiation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008985; ConA like lec_gl.
DR InterPro; IPR006558; LamG like.
DR InterPro; IPR008000; Notch dom.
DR InterPro; IPR008754; Peptidase_M43B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF05572; Peptidase_M46; I.
DR Pfam; PF00084; sushi; 5.
DR SMART; SM00032; CCP; 4.
DR SMART; SM00560; LamGL; 1.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 1574 AA; 176105 MW; AE53F0187E6EFB7B CRC64;

Query Match 92.1%; Score 35; DB 11; Length 1574;
Best Local Similarity 83.3%; Pred. No. 71;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 1357 CVPVTC 1362

RESULT 15
Q7Z408 PRELIMINARY; PRT; 3487 AA.
ID Q7Z408
AC Q7Z408

```

DT 01-OCT-2003 (TremBurel. 25, Created)  
 DT 01-OCT-2003 (TremBurel. 25, Last sequence update)  
 DT 01-OCT-2003 (TremBurel. 25, Last annotation update)  
 DE CUB and sushi multiple domains 2.  
 GN CSMD2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22788796; PubMed=12906867;  
 RA Lau W L.; Scholnick S.B.;  
 RT "Identification of two new members of the CSMD gene family."  
 RL Genomics 82:412-415(2003).  
 DR EMBL; AY210418; ARO34701.1; -.  
 SQ SEQUENCE 3487 AA; 380035 MW; 3A77D8457DF8EFB5 CRC64;  
  
 Query Match 92.1%; Score 35; DB 4; Length 3487;  
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 Qy 1 CVPLTC 6  
 Db 2588 CVPVTC 2593

Search completed: September 5, 2004, 10:00:05  
 Job time : 15.5758 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:37:49 ; Search time 19.697 Seconds  
(without alignments)  
86.068 Million cell updates/sec

Title: US-09-761-636A-12

Perfect score: 38

Sequence: 1 CVPLTC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	6	4	Aau04531
2	36	94.7	144	5	ABBS3096
3	36	94.7	613	4	Aau00816
4	36	94.7	882	4	AAB83372
5	36	94.7	883	4	AAB83371
6	36	94.7	3069	5	AAE20787
7	36	94.7	3069	5	AAE20900
8	36	94.7	3095	5	AAE20788
9	36	94.7	3100	5	AAE20789
10	36	94.7	3100	5	AAE20789
11	36	94.7	3104	5	ABG79168
12	35	92.1	95	4	AAU48139
13	35	92.1	95	6	ABM44718
14	35	92.1	310	6	ADA55489
15	35	92.1	368	7	ADD27716
16	35	92.1	530	7	ADD46650
17	35	92.1	530	7	ADBD4650
18	35	92.1	1603	4	ABG55158
19	35	92.1	1752	4	ABG11138
20	34	89.5	570	7	ABG65688
21	33	86.8	10	3	AAU40081
22	33	86.8	60	4	AAU40081
23	33	86.8	60	4	AAU66225
24	33	86.8	60	6	ABM36600
25	33	86.8	60	6	ABM62744

26	33	86.8	66	5	ABP29076
27	33	86.8	138	4	AAU67640
28	33	86.8	138	6	ABM64159
29	33	86.8	158	6	ABM64758
30	33	86.8	158	6	ABU16591
31	33	86.8	159	6	ADA33421
32	33	86.8	194	4	AAU58878
33	33	86.8	194	6	ABM52397
34	33	86.8	556	6	ABM52397
35	33	86.8	637	2	AAU25004
36	33	86.8	688	1	AAU80361
37	33	86.8	1236	4	ABM60480
38	33	86.8	1524	2	AAU15458
39	33	86.8	1575	6	ABR58468
40	33	86.8	1587	3	AAAB40917
41	33	86.8	1587	5	AAAM50361
42	33	86.8	1587	6	ABR58467
43	32	84.2	48	6	ADA57157
44	32	84.2	48	6	ADA41023
45	32	84.2	49	3	AAU91362

ALIGNMENTS

RESULT 1  
AAU04531  
ID AAU04531 standard; peptide; 6 AA.  
XX AC AAU04531;  
XX XX  
DT 26-SEP-2001 (first entry)  
XX XX  
DE VEGF based monocyclic peptide 9.  
XX XX  
KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
KW diabetic retinopathy; chronic inflammation; cyclic.  
XX OS Synthetic.  
XX XX  
PH Key Location/Qualifiers  
FT Disulfide-bond 1..6  
FT /note= "This bond cyclises the peptide"  
XX XX  
PN WO200152875-A1.  
XX XX  
PD 26-JUL-2001.  
XX XX  
PF 18-JAN-2001; 2001WO-US001533.  
XX XX  
PR 18-JAN-2000; 2000US-0176293P.  
PR 16-MAY-2000; 2000US-0204590P.  
XX XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX XX  
PI Achen MG, Hughes RA, Stacker S, Cendron A;  
XX WPI; 2001-442248/47.  
XX XX  
PT Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
PT or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
PT from an exposed loop of a growth factor protein by oxidizing the cysteine  
PT residues.  
XX XX  
PS Claim 49; Page 32; 102pp; English.  
XX XX  
CC The sequence represents a monomeric monocyclic peptide of the invention,  
CC whose 3-dimensional structure is modelled on the exposed loop of human  
CC VEGF (vascular endothelial growth factor). The invention relates to a  
CC method of producing a monomeric monocyclic peptide by a measuring beta-  
CC beta carbon separation distances on opposite antiparallel strands of a

CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis,  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 SQ Sequence 6 AA;

Query Match 100.0%; Score 38; DB 4; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6  
 |||||  
 Db 1 CVPLTC 6

RESULT 2  
 ABB53096  
 ID ABB53096 standard; protein; 144 AA.

AC ABB53096;  
 DT 12-FEB-2002 (first entry)  
 XX Human ORF2 protein.  
 DE  
 KW OREX; human; pharmacogenomics; cancer; hyperproliferative disorder;  
 KW dysproliferative disorder; neurodegenerative disorder; organ transplant;  
 KW cardiovascular disease; cytokine; cell proliferation; immunomodulatory;  
 KW cell differentiation; haematopoiesis; tissue growth; thrombolytic;  
 KW neutropic; neuroprotective; antiarthritic; antimicrobial; vulnery;  
 KW cytostatic; antidiabetic; virucide; antiinfertility; anticonvulsant;  
 KW vasotropic; antiparkinsonian; dermatological; antineumatic; antitumor;  
 KW antitumor; osteopathic; tranquiliser; cerebroprotective; hepatotropic;  
 KW antiinflammatory; tumour inhibition.

OS Homo sapiens.

FN WO200177155-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-US040469.

XX 06-APR-2000; 2000US-0195576P.

PR 05-APR-2001; 2001US-0082673A.

XX (CURA-) CURAGEN CORP.

XX Fernandes E, Vernet CAM, Mishnu VS, Leach MD, Shimkets RA;

PI Zerhusen BD, Kekuda R;

XX WPI; 2002-049131/06.

DR N-PSDB; ABA90195.

PT New polypeptides and polynucleotides, useful for treating diseases such  
 PT as cancer, Alzheimer's disease, atherosclerosis, diseases associated with  
 PT liver, comprises polypeptides and polynucleotides of open reading frame  
 PT proteins.

XX Claim 11; Page 82; 11pp; English.

XX The present invention provides the protein and coding sequences of human  
 CC ORFX, where X is any number between 1 and 132. These sequences can be  
 CC used to diagnose and treat ORFX related disorders, including cancer,  
 CC hyperproliferative and dysproliferative disorders, neurodegenerative  
 CC disorders, disorders related to organ transplantation or cardiovascular  
 CC diseases, and may have cell proliferation or differentiation, cytokine,  
 CC immunomodulatory, haematopoiesis regulating, tissue growth, inhibin or  
 CC activin, chemotactic or chemokinetic, haemostatic or thrombolytic and/or  
 CC antiinflammatory activities. The present sequence is one of the ORFX  
 CC proteins of the invention

XX Sequence 144 AA;

Query Match 94.7%; Score 36; DB 5; Length 144;  
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6  
 |||||  
 Db 117 CVPLTC 122

RESULT 3  
 AAU00816  
 ID AAU00816 standard; protein; 613 AA.

AC AAU00816;

XX 01-JUN-2001 (first entry)

XX Human Immunoglobulin superfamily, IgSF, protein #2.

XX Human; Immunoglobulin superfamily protein; IgSF; Immune response;  
 KW inflammatory response; cell-cell interaction; cell-surface recognition;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; pulmonary disorder;  
 KW cardiovascular disorder; renal disorder; proliferative disorder; cancer;  
 KW common variable adhesion deficiency syndrome; AIDS; SCID;  
 KW acquired immunodeficiency syndrome; anaemia; rheumatoid arthritis;  
 KW Alzheimer's disease; Crohn's disease; liver cancer; leukaemia;  
 KW Hodgkin's lymphoma; Parkinson's disease; Huntington's disease; dementia;  
 KW atherosclerosis; stroke; diabetes mellitus; Addison's disease;  
 KW urticaria; severe combined immunodeficiency; antibody.

XX Homo sapiens.

OS Key Location/Qualifiers

FH Domain 4..52 /label= Immunoglobulin\_like\_domain

FT Region 6..11 /label= Immunogenic\_epitope

FT Region 73..79 /label= Immunogenic\_epitope

FT Region 121..127 /label= Immunogenic\_epitope

FT Region 141..146 /label= Immunogenic\_epitope

FT Region 223..230 /label= Immunogenic\_epitope

FT Region 253..260 /label= Immunogenic\_epitope

FT Region 328..336 /label= Immunogenic\_epitope

FT Region 374..383 /label= Immunogenic\_epitope

FT Region 392..404 /label= Immunogenic\_epitope

FT Region /label= Immunogenic\_epitope  
 484..500  
 FT Region /label= Immunogenic\_epitope  
 511..517  
 FT Region /label= Immunogenic\_epitope  
 577..583  
 FT Region /label= Immunogenic\_epitope  
 605..610  
 FT Region /label= Immunogenic\_epitope  
 WO200118176-A1.  
 XX PN  
 XX XX  
 XX PD 15-MAR-2001.  
 XX XX  
 XX PF 29-AUG-2000; 2000WO-US023662.  
 XX XX  
 XX PR 03-SEP-1999; 99US-0152248P.  
 XX XX  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PA (NIJ)/ NI J.  
 XX XX  
 XX XX Young PE, Ruben SM, Shi Y;  
 XX PI  
 XX XX WPI; 2001-203084/20.  
 DR DR N-PSDB; AAS00146.  
 XX XX  
 XX XX Isolated nucleic acid molecule encoding a human secreted protein is used  
 FT FT in preventing, treating or ameliorating a medical condition.  
 XX XX  
 XX PS Claim 11; Page 238-239; 247pp; English.  
 XX XX  
 CC The sequence is a Human Immunoglobulin superfamily, IgSF, protein, a  
 CC diverse family of proteins involved in cell-cell interactions, cell-  
 CC surface recognition, intercellular communication and immune and  
 CC inflammatory responses. Polypeptides and antibodies directed to  
 CC polypeptides of the present invention are useful to provide immunological  
 CC probes for differential identification of tissues. Antibodies can be used  
 CC to assay levels of polypeptides encoded by polynucleotides of the  
 CC invention. Polypeptides of the present invention can be used to treat or  
 CC prevent diseases or conditions such as neural disorders, immune system  
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
 CC disorders, pulmonary disorders, cardiovascular disorders, renal  
 CC disorders, proliferative disorders, and/or cancerous diseases and  
 CC conditions. Polynucleotides of the invention are also useful in treating  
 CC the above disorders. Examples of the disorders include common variable  
 CC adhesion deficiency syndrome, acquired immunodeficiency syndrome (AIDS),  
 CC anaemia, rheumatoid arthritis, Alzheimer's disease, Crohn's disease,  
 CC liver cancer, leukaemia, Hodgkin's lymphoma, Parkinson's disease,  
 CC Huntington's disease, dementia, arteriosclerosis, stroke, diabetes  
 CC mellitus, Addison's disease, urticaria, severe combined immunodeficiency  
 CC (SCID). Many more examples of diseases and disorders are given in the  
 CC specification  
 XX XX  
 SQ Sequence 613 AA;  
 Query Match 94.7%; Score 36; DB 4; Length 613;  
 Best Local Similarity 83.3%; Pred. No. 6.7e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVP LTC 6  
 |||:|  
 Db 264 CVP LTC 269  
 RESULT 4  
 AAB83372  
 ID AAB83372 standard; protein; 882 AA.  
 XX AC AAB83372;  
 XX XX  
 XX DT 26-MAR-2002 (first entry)  
 XX DE NOV16 protein sequence.

XX NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;  
 KW cardiovascular; casein kinase II phosphorylation site; contraception;  
 KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;  
 KW epidermal growth factor; cell development; apoptosis; cell adhesion;  
 KW growth migration; cell structure; motility; cancer; immune disorder;  
 KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;  
 KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;  
 KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;  
 KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.  
 XX Unidentified.  
 XX OS  
 XX WO200136638-A2.  
 PN  
 XX PD 25-MAY-2001.  
 XX XX  
 XX PF 17-NOV-2000; 2000WO-US031543.  
 XX XX  
 XX PR 19-NOV-1999; 99US-0166336P.  
 XX PR 29-NOV-1999; 99US-0167785P.  
 XX PR 08-MAR-2000; 2000US-0187844P.  
 XX PR 16-NOV-2000; 2000US-00715417.  
 XX XX  
 XX PA (CURA-) CURAGEN CORP.  
 XX XX  
 XX PI Shimkets RA, Lichenstein H, Vernet C, Fernandes E;  
 XX WPI; 2001-648134/74.  
 DR DR N-PSDB; AAF87127.  
 DR DR  
 XX PT Novel human polypeptides and the nucleic acids that encode them useful  
 PT for preventing, diagnosing and treating e.g. cancer, inflammation and  
 PT immune disorders.  
 XX XX  
 PS Claim 1; Page 50-52; 141pp; English.  
 XX XX  
 CC This sequence is the NOV16 protein. The invention relates to the NOV1-  
 CC NOV16 proteins, and their coding sequences. The proteins have Cytostatic;  
 CC contraceptive; antiinflammatory; immunomodulatory; and cardiovascular  
 CC activities. The sequences may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate NOVX expression. They  
 CC may be used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of protein by expressing inactive proteins or to supplement the  
 CC patients own production of protein. They are used to produce NOVX  
 CC proteins, by inserting the nucleic acid into a cell and culturing it to  
 CC express the protein. The DNA may be used as DNA probes in assays to  
 CC detect and quantitate the presence of similar DNAs in samples, and which  
 CC patients may need restorative therapy. The NOVX protein may also be used  
 CC as antigens in the production of antibodies (Abs) against NOVX and in  
 CC assays to identify modulators of NOVX expression and activity. The anti-  
 CC NOVX Abs and antagonist are used to down regulate expression and  
 CC activity. The anti-NOVX Abs are used for detecting the presence of NOVX  
 CC in samples. Disorders that may be prevented, diagnosed and/or treated  
 CC vary depending on the NOVX protein. NOV1, NOV3, NOV5, NOV7, NOV9-11 and  
 CC NOV13-16 have casein kinase II phosphorylation sites characteristic of  
 CC serine/threonine kinases, and are used to treat kinase-related disorders  
 CC (e.g. Peutz-Jeghers syndrome, cellular proliferation and contraception).  
 CC NOV2-3, NOV6 and NOV8 are homologous to the epidermal growth factor (EGF)  
 CC -like super family and are involved in, e.g. regulation of cell  
 CC development, apoptosis, cell adhesion, growth migration, cell structure  
 CC and motility and protein management, and are used to treat cancers,  
 CC inflammatory disorders, immune disorders and cellular adhesion disorders.  
 CC NOV6-10 are homologous to EGF-like fibrillin proteins and are used to  
 CC treat cardiovascular disease e.g. hypertrophic cardiomyopathy, long-QT  
 CC syndrome and marfan syndrome  
 XX XX  
 SQ Sequence 882 AA;  
 Query Match 94.7%; Score 36; DB 4; Length 882;  
 Best Local Similarity 83.3%; Pred. No. 9.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CC (e.g. Peutz-Jeghers syndrome, cellular proliferation and contraception).  
 CC NOV2-3, NOV6 and NOV8 are homologous to the epidermal growth factor (EGF)  
 CC -like super family and are involved in, e.g. regulation of cell  
 CC development, apoptosis, cell adhesion, growth migration, cell structure  
 CC and motility and protein management, and are used to treat cancers.  
 CC inflammatory disorders, immune disorders and cellular adhesion disorders.  
 CC NOV6-10 are homologous to EGF-like fibrillin proteins and are used to  
 CC treat cardiovascular disease e.g. hypertrophic cardiomyopathy, long-QT  
 CC syndrome and marfan syndrome  
 CC  
 XX Sequence 883 AA;  
 SQ

Query Match 94.7%; Score 36; DB 4; Length 883;  
 Best Local Similarity 83.3%; Pred. No. 9.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPITC 6  
 Db 264 CVPITC 269  
 |||:|

RESULT 6  
 AAE20787  
 ID AAE20787 standard; protein; 3069 AA.  
 XX  
 AC AAE20787;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Human C3b/C4b complement receptor like protein #1.  
 XX  
 KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;  
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;  
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;  
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;  
 KW transplant rejection; autoimmune disease; ischaemic condition; neotropic;  
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;  
 KW infertility; vasodilator; obesity; cardiac.  
 XX  
 OS Homo sapiens.  
 XX

Key Location/Qualifiers  
 FH Misc-difference 387  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 461  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 586  
 FT Misc-difference /label= Unknown  
 FT Misc-difference 1272  
 FT Misc-difference /label= Unknown  
 XX  
 XX WO200210199-A2.  
 XX  
 XX 07-FEB-2002.  
 XX  
 XX 24-JUL-2001; 2001WO-US023232.  
 XX  
 XX 02-AUG-2000; 2000US-0222504P.  
 XX 28-NOV-2000; 2000US-00728787.  
 XX  
 XX (AMGE-) AMGEN INC.  
 XX  
 XX Welcher AA, Elliott GS;  
 XX  
 XX WPI; 2002-303934/34.  
 XX N-PSDB; AAD33318.  
 XX  
 XX Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic  
 XX acid molecule, useful for treating, preventing and diagnosing rheumatoid  
 XX arthritis, psoriatic arthritis, inflammatory arthritis, and multiple  
 XX sclerosis.  
 XX  
 XX Claim 13; Fig 1; 251pp; English.

QY 1 CVPITC 6  
 Db 264 CVPITC 269  
 |||:|

RESULT 5  
 AAB83371  
 ID AAB83371 standard; protein; 883 AA.  
 XX  
 AC AAB83371;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE NOV15 protein sequence.  
 XX  
 XX NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;  
 KW cardiovascular; casein kinase II phosphorylation site; contraception;  
 KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;  
 KW epidermal growth factor; cell development; apoptosis; cell adhesion;  
 KW growth migration; cell structure; motility; cancer; immune disorder;  
 KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;  
 KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;  
 KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;  
 KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.  
 XX  
 XX Unidentified.  
 OS  
 XX  
 XX WO200136638-A2.  
 XX  
 PD 25-MAY-2001.  
 XX  
 XX 17-NOV-2000; 2000WO-US031543.  
 XX  
 XX 19-NOV-1999; 99US-0166336P.  
 XX 29-NOV-1999; 99US-0167785P.  
 XX 08-MAR-2000; 2000US-0187844P.  
 XX 16-NOV-2000; 2000US-00715417.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Lichenstein H, Vernet C, Fernandes E;  
 XX  
 XX WPI; 2001-648134/74.  
 XX N-PSDB; AAF87126.  
 XX  
 XX Novel human polypeptides and the nucleic acids that encode them useful  
 XX for preventing, diagnosing and treating e.g. cancer, inflammation and  
 XX immune disorders.  
 XX  
 XX Claim 1; Page 46-48; 141pp; English.

This sequence is the NOV15 protein. The invention relates to the NOV1-  
 NOV16 proteins, and their coding sequences. The proteins have Cytostatic;  
 contraceptive; antiinflammatory; immunomodulatory; and cardiovascular  
 activities. The sequences may be used in the prevention, diagnosis and  
 treatment of diseases associated with inappropriate NOVX expression. They  
 may be used to treat disorders associated with decreased expression by  
 rectifying mutations or deletions in a patient's genome that affect the  
 activity of protein by expressing inactive proteins or to supplement the  
 patients own production of protein. They are used to produce NOVX  
 proteins, by inserting the nucleic acid into a cell and culturing it to  
 express the protein. The DNA may be used as DNA probes in assays to  
 detect and quantitate the presence of similar DNAs in samples, and which  
 patients may need restorative therapy. The NOVX protein may also be used  
 as antigens in the production of antibodies (Abs) against NOVX and in  
 assays to identify modulators of NOVX expression and activity. The anti-  
 NOVX Abs and antagonist are used to down regulate expression and  
 activity. The anti-NOVX Abs are used for detecting the presence of NOVX  
 in samples. Disorders that may be prevented, diagnosed and/or treated  
 vary depending on the NOVX protein. NOV1, NOV3, NOV5, NOV7, NOV9-11 and  
 NOV13-16 have casein kinase II phosphorylation sites characteristic of  
 serine/threonine kinases, and are used to treat kinase-related disorders



Welcher AA, Elliott GS;

28-NOV-2000; 2000US-00/28787;  
XX



XX WO200210199-A2.  
 XX 07-FEB-2002.  
 XX 24-JUL-2001; 2001WO-US023232.  
 XX 02-AUG-2000; 2000US-0222504P.  
 XX 28-NOV-2000; 2000US-00728787.  
 XX (AMGE-) AMGEN INC.  
 XX Welcher AA, Elliott GS;  
 XX WPI; 2002-303934/34.  
 XX N-PSDB; AAD33320.  
 XX Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic  
 XX acid molecule, useful for treating, preventing and diagnosing rheumatoid  
 XX arthritis, psoriatic arthritis, inflammatory arthritis, and multiple  
 XX sclerosis.  
 XX Claim 13; Page 239-251; 251pp; English.  
 XX The invention relates to a nucleic acid encoding a novel C3b/C4b  
 XX complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like  
 XX polypeptide and nucleic acid molecules may be used to treat, prevent,  
 XX ameliorate, diagnose and/or detect diseases such as immune system  
 XX disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory  
 XX arthritis, osteoarthritis, inflammatory joint disease, autoimmune  
 XX disease, multiple sclerosis, lupus, inflammatory bowel disease,  
 XX transplant rejection, nervous system disorders (e.g. Alzheimer's  
 XX disease), ischaemic conditions, metabolic disorders (e.g. obesity and  
 XX diabetes) and infertility. The invention is useful in gene therapy. The  
 XX present sequence is human C3b/C4b complement receptor like protein,  
 XX alternative version. Note: The present sequence is stated to be the same  
 XX as that referred to as SEQ ID NO:7 (AA20789) shown in figure 2 of the  
 XX specification. However the sequences differ at position 726  
 XX Sequence 3100 AA;  
 XX  
 XX Query Match 94.7%; Score 36; DB 5; Length 3100;  
 XX Best Local Similarity 83.3%; Pred. No. 2.8e+03;  
 XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVP LTC 6  
 |||:|  
 Db 2269 CVP LTC 2274  
 XX  
 XX RESULT 11  
 XX ABG79168  
 XX ID ABG79168 standard; protein; 3104 AA.  
 XX AC ABG79168;  
 XX DT 15-NOV-2002 (first entry)  
 XX Human cub and sushi domain containing protein #1.  
 XX Human; NOVX; neurodegenerative disease; Alzheimer's disease; anxiety;  
 XX Parkinson's disease; Huntington's disease; neurological disorder;  
 XX schizophrenia; manic depression; mental retardation; angina pectoris;  
 XX cardiovascular disease; acute heart failure; myocardial infarction;  
 XX muscular disease; muscular disorder; retinal disease; photoreception;  
 XX deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;  
 XX immunological disorder; inflammatory disease; immune disease; diabetes;  
 XX bacterial infection; fungal infection; protozoal infection; obesity;  
 XX viral infection; reproductive system disorder; metabolic disturbance;  
 XX anorexia; wasting disorder; chronic disease; infectious disease;  
 XX dyslipidaemia; cub; sushi; myelin; von willebrand factor; kielin;  
 XX semaphorin; serine/threonine protein kinase; TGF-beta binding;  
 XX mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;

KW tolloid-like 2; cysteine sulfinic acid decarboxylase SNP;  
 KW single nucleotide polymorphism.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX misc\_difference 1027  
 XX /note= "Ala substituted by Thr as a result of a single  
 XX nucleotide polymorphism (SNP)"  
 XX misc\_difference 1084  
 XX /note= "Val substituted by Ala as a result of a single  
 XX nucleotide polymorphism (SNP)"  
 XX misc\_difference 1362  
 XX /note= "Asp substituted by Gly as a result of a single  
 XX nucleotide polymorphism (SNP)"  
 XX WO200264791-A2.  
 XX 22-AUG-2002.  
 XX 10-DEC-2001; 2001WO-US048369.  
 XX 08-DEC-2000; 2000US-0254329P.  
 XX 14-DEC-2000; 2000US-0255648P.  
 XX 15-MAY-2001; 2001US-0291037P.  
 XX 08-JUN-2001; 2001US-0297173P.  
 XX 08-JUN-2001; 2001US-0309258P.  
 XX 29-AUG-2001; 2001US-0315639P.  
 XX 01-OCT-2001; 2001US-0326393P.  
 XX (CURA-) CURAGEN CORP.  
 XX Alsbrook JP, Anderson DW, Burgess CE, Boldog FL, Casman SU;  
 XX Colman SD, Edinger SR, Ellerman K, Gerlach V, Gorman L, Grosse WM;  
 XX Guo X, Herrmann JL, Kekuda R, Lepley DM, Li L, MacDougall JR;  
 XX Millett I, Pena CA, Peyman JA, Rastelli L, Rieger DK, Shimkets RA;  
 XX Smithson G, Spytek KA, Stone DJ, Tchernev VT, Vernet CAM, Voss EZ;  
 XX Zerhusen BD, Zhong H, Zhong M;  
 XX WPI; 2002-643486/69.  
 XX N-PSDB; ABS64375.  
 XX New NOVX polypeptides and polynucleotides useful for treating or  
 XX preventing e.g. neurodegenerative diseases, neurological disorders,  
 XX cardiovascular diseases, muscular diseases and disorders, or  
 XX immunological diseases.  
 XX Claim 1; Page 13-14; 299pp; English.  
 XX The present invention relates to new NOVX polypeptides. The polypeptides,  
 XX polynucleotides and antibodies are useful in the manufacture of a  
 XX medicament for treating or preventing neurodegenerative diseases (e.g.  
 XX Alzheimer's disease, Parkinson's disease, or Huntington's disease).  
 XX neurological disorders (e.g. anxiety, schizophrenia, manic depression or  
 XX mental retardation), cardiovascular disease (e.g. acute heart failure,  
 XX angina pectoris or myocardial infarction), muscular diseases and  
 XX disorders, retinal diseases (including those involving photoreception,  
 XX deafness and keratinisation disorders), cancer (e.g. ovarian cancer or  
 XX melanoma), immunological disorders, inflammatory and immune diseases,  
 XX bacterial, fungal, protozoal and viral infections, and reproductive  
 XX system disorders. The proteins of the invention may be used to screen  
 XX drugs or compounds that modulate the NOVX protein activity or expression,  
 XX as well as to treat disorders characterised by insufficient or excessive  
 XX production of NOVX protein or protein forms that have decreased or  
 XX aberrant activity compared to NOVX wild type protein, such as diabetes,  
 XX obesity, metabolic disturbances associated with obesity, anorexia and  
 XX wasting disorders associated with chronic diseases and various cancers,  
 XX infectious diseases and various dyslipidaemias. The nucleic acid  
 XX sequences of the invention may be used in chromosome mapping, identifying  
 XX an individual from minute biological samples (tissue typing), and in  
 XX forensic identification of a biological sample. The present amino acid  
 XX sequence represents a NOVX protein of the invention

SQ Sequence 3104 AA;  
 Query Match 94.7%; Score 36; DB 5; Length 3104;  
 Best Local Similarity 83.3%; Pred. No. 2.8e+03;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPITC 6  
 |||:  
 Db 2486 CVPITC 2491

RESULT 12  
 AAU48199  
 ID AAU48199 standard; protein; 95 AA.  
 XX  
 AC AAU48199;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #9095.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US012865.  
 XX  
 PR 21-APR-2000; 2000US-0199047P.  
 PR 02-JUN-2000; 2000US-0208841P.  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI: 2001-616774/71.  
 DR N-PSDB; AAS95942.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 XX  
 PS Example 1; SEQ ID NO 9394; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 95 AA;

Query Match 92.1%; Score 35; DB 6; Length 95;  
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 Db 62 CVPVTC 67  
 |||:|

RESULT 14  
 ADA55489  
 ID ADA55489 standard; protein; 310 AA.

AC ADA55489;

DT 20-NOV-2003 (first entry)

DE Human protein, SEQ ID 3057.

KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
 KW Gene therapy; human; secretory protein; membrane proteins; cancer;  
 KW inflammatory disease; osteoporosis; neurological disease.

XX Homo sapiens.

XX EPI293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-00328381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masubo Y;

XX WPI; 2003-395539/38.

XX N-PSDB; ADA53850.

XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
 PT and/or membrane proteins, useful for developing medicines for diseases in  
 PT which the gene is involved, or as target molecules for gene therapy.

PS Claim 14; SEQ ID NO 3057; 205pp; English.

XX The present invention relates to novel human secretory or membrane  
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
 CC ADA54071). The coding sequences are useful in the gene therapy of  
 CC diseases caused by abnormalities of the proteins, e.g. cancer,  
 CC inflammatory diseases, osteoporosis or neurological disease.

XX Sequence 310 AA;

Query Match 92.1%; Score 35; DB 6; Length 310;  
 Best Local Similarity 83.3%; Pred. No. 5.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 Db 262 CLPLTC 267  
 |:|

RESULT 15

ADD27716

ID ADD27716 standard; protein; 368 AA.

XX AC ADD27716;

XX 15-JAN-2004 (first entry)

XX Human adipocyte Selected Interacting Domain (SID) prey protein #1173.

KW Adipocyte; protein-protein interaction; protein complex;  
 KW bait-prey complex; Selected Interacting Domain; SID; drug screening;  
 KW drug discovery; metabolic disease; obesity; lipodystrophy;  
 KW diabetes mellitus; type 2; non-insulin dependent; NIDDM;  
 KW adipogenesis modulation; gene therapy; human.

XX Homo sapiens.

XX WO200290544-A2.

XX 14-NOV-2002.

XX 03-MAY-2002; 2002WO-EP006333.

XX 04-MAY-2001; 2001US-0288885P.

XX (HYBR-) HYBRIGENICS.

XX (LYNX-) LYNX THERAPEUTICS INC.

XX Legrain P, Whiteside S, Mao J, Khrebtukova I, Luo S;  
 PI WPI; 2003-111975/10.

XX N-PSDB; ADD27715.

XX New complex of bait and prey between two polypeptides or polynucleotides  
 PT encoding the two polypeptides of adipocytes, useful for selecting a  
 PT modulating compound that inhibits or activates protein-protein  
 PT interactions.

PS Claim 6; SEQ ID NO 1173; 232pp; English.

XX The invention relates to a bait-prey complex between two adipocyte  
 CC polypeptides, or between two polynucleotides encoding adipocyte  
 CC polypeptides. The invention also relates to selected interacting Domain  
 CC (SID) polypeptides which interact with selected bait polypeptides;  
 CC polynucleotides encoding SID polypeptides; vectors comprising SID  
 CC polynucleotides; recombinant host cells comprising an adipocyte  
 CC polynucleotide or a SID-encoding vector; a method of selecting for a  
 CC compound which modulates interactions between adipocyte polypeptides;  
 CC adipocyte modulator compounds identified using the method; a  
 CC pharmaceutical composition comprising an adipocyte modulator, or a SID-  
 CC encoding vector or host cell; and a protein chip comprising adipocyte  
 CC bait polypeptides. The bait-prey complexes of the invention are useful  
 CC for selecting a compound that inhibits or activates protein-protein  
 CC interactions between adipocyte polypeptides. The modulatory compounds  
 CC identified can be used in the treatment of metabolic diseases such as  
 CC obesity, lipodystrophy and type 2 diabetes mellitus, and in the  
 CC modulation of adipogenesis. The present sequence represents a  
 CC specifically claimed adipocyte SID prey polypeptide of the invention.

XX Sequence 368 AA;

Query Match 92.1%; Score 35; DB 7; Length 368;  
 Best Local Similarity 83.3%; Pred. No. 6.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
 Db 151 CVPVTC 156  
 |||:|

Search completed: September 5, 2004, 09:55:19  
 Job time : 21.697 secs

us-09-761-636a-12.open.rag

Sun Sep 5 10:36:20 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 10:00:15 ; Search time 16.8485 Seconds  
(without alignments)  
112.199 Million cell updates/sec

Title: US-09-761-636A-12

Perfect score: 38

Sequence: 1 CVPLTC 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT NEW PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	6	9	US-09-761-636A-12
2	38	100.0	45	12	US-10-424-599-164353
3	36	94.7	175	12	US-09-826-734-122
4	36	94.7	613	9	US-09-799-514-9
5	36	94.7	3104	12	US-10-016-248-2
6	36	94.7	3389	12	US-10-016-248-47
7	36	94.7	3389	16	US-10-408-765A-2286
8	36	94.7	3508	12	US-10-016-248-46
9	36	94.7	3564	12	US-10-016-248-45
10	35	92.1	127	12	US-10-424-599-198429
11	35	92.1	310	15	US-10-094-749-3057
12	35	92.1	1627	10	US-09-983-025-25
13	35	92.1	1627	15	US-10-295-027-663
14	34	89.5	80	12	US-10-424-599-272858
15	34	89.5	570	15	US-10-104-047-3842

16	33	86.8	52	16	US-10-437-963-190243	Sequence 130243,
17	33	86.8	58	12	US-10-424-599-252356	Sequence 252356,
18	33	86.8	60	12	US-10-424-599-198470	Sequence 198470,
19	33	86.8	62	12	US-10-424-599-170802	Sequence 170802,
20	33	86.8	69	12	US-10-424-599-145317	Sequence 145317,
21	33	86.8	103	12	US-10-424-599-217717	Sequence 217717,
22	33	86.8	149	12	US-10-425-114-60690	Sequence 60690, A
23	33	86.8	158	12	US-10-282-122A-44515	Sequence 44515, A
24	33	86.8	162	12	US-10-425-114-50640	Sequence 50640, A
25	33	86.8	187	12	US-10-425-114-50780	Sequence 50780, A
26	33	86.8	556	9	US-09-815-923-6	Sequence 6, Appli
27	33	86.8	1575	12	US-10-262-839-212	Sequence 212, App
28	33	86.8	1587	9	US-09-845-583-10	Sequence 10, Appl
29	33	86.8	1587	12	US-10-262-839-210	Sequence 210, App
30	32	84.2	34	16	US-10-437-963-200127	Sequence 200127,
31	32	84.2	49	9	US-09-739-254-83	Sequence 83, Appl
32	32	84.2	49	9	US-09-904-615-83	Sequence 83, Appl
33	32	84.2	49	14	US-10-054-988-83	Sequence 83, Appl
34	32	84.2	49	14	US-10-055-098-83	Sequence 83, Appl
35	32	84.2	55	12	US-10-424-599-207242	Sequence 207242,
36	32	84.2	57	12	US-10-424-599-240922	Sequence 240922,
37	32	84.2	117	12	US-10-424-599-274520	Sequence 274520,
38	32	84.2	125	9	US-09-864-761-46679	Sequence 46679, A
39	32	84.2	164	12	US-10-424-599-162389	Sequence 162389,
40	32	84.2	235	14	US-10-106-698-6300	Sequence 6300, Ap
41	32	84.2	256	12	US-10-466-164-65	Sequence 65, Appl
42	32	84.2	289	14	US-10-017-161-2008	Sequence 2008, Ap
43	32	84.2	289	15	US-10-292-798-1654	Sequence 1654, Ap
44	32	84.2	344	9	US-09-898-570-18	Sequence 18, Appl
45	32	84.2	344	10	US-09-839-446-18	Sequence 18, Appl

## ALIGNMENTS

## RESULT 1

US-09-761-636A-12  
; Sequence 12, Application US/09761636A  
; Patent No. US20020065218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09/761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-761-636A-12

Query Match 100.0%; Score 38; DB 9; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVPLTC 6

Db 1 CVPLTC 6

## RESULT 2

US-10-424-599-164353  
; Sequence 164353, Application US/10424599  
; Publication No. US20040031072A1

Sun Sep 5 10:36:21 2004

us-09-761-636a-12.open.rapb

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 164353
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_119429C.1.pep
US-10-424-599-164353

Query Match          100.0%; Score 38; DB 12; Length 45;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 34 CVPLTC 39

RESULT 3
US-09-826-734-122
; Sequence 122, Application US/09826734
; Publication No. US20030017457A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Mishra, Vishnu S.
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Zethusen, Bryan D.
; APPLICANT: Kekuda, Ramesha
; TITLE OF INVENTION: Novel Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-754
; CURRENT APPLICATION NUMBER: US/09/826,734
; CURRENT FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/195,576
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 270
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 122
; LENGTH: 175
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-826-734-122

Query Match          94.7%; Score 36; DB 12; Length 175;
Best Local Similarity 83.3%; Pred. No. 96;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 117 CVPLTC 122

RESULT 4
US-09-799-514-9
; Sequence 9, Application US/09799514
; Patent No. US20020065220A1
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: Immunoglobulin Superfamily Polynucleotides, Polypeptides, and Ap
; FILE REFERENCE: PT015P1
; CURRENT APPLICATION NUMBER: US/09/799,514
; CURRENT FILING DATE: 2001-03-07
```

```
; PRIOR APPLICATION NUMBER: PCT/US00/23662
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/152,248
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-799-514-9

Query Match          94.7%; Score 36; DB 9; Length 613;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 264 CVPLTC 269

RESULT 5
US-10-016-248-2
; Sequence 2, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3104
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-248-2

Query Match          94.7%; Score 36; DB 12; Length 3104;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6
Db 2486 CVPLTC 2491

RESULT 6
US-10-016-248-47
; Sequence 47, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
```



```

; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 3389
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-248-47

```

```

Query Match          94.7%; Score 36; DB 12; Length 3389;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CVP LTC 6
Db      2677 CVP LTC 2682

```

## RESULT 7

```

US-10-408-765A-2286
; Sequence 2286, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2286
; LENGTH: 3389
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2286

```

```

Query Match          94.7%; Score 36; DB 16; Length 3389;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CVP LTC 6
Db      2677 CVP LTC 2682

```

## RESULT 8

```

US-10-016-248-46
; Sequence 46, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248

```

```

; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 3508
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-016-248-46

```

```

Query Match          94.7%; Score 36; DB 12; Length 3508;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CVP LTC 6
Db      2677 CVP LTC 2682

```

## RESULT 9

```

US-10-016-248-45
; Sequence 45, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 3564
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-016-248-45

```

```

Query Match          94.7%; Score 36; DB 12; Length 3564;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CVP LTC 6
Db      2733 CVP LTC 2738

```

## RESULT 10

US-10-424-599-198429  
; Sequence 198429, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 198429  
; LENGTH: 127  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(127)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_21205C.1.pap  
US-10-424-599-198429

Query Match 92.1%; Score 35; DB 12; Length 127;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6

Db 116 CLPLTC 121

## RESULT 11

US-10-094-749-3057  
; Sequence 3057, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHITO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3057  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-094-749-3057

Query Match 92.1%; Score 35; DB 15; Length 310;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6

Db 262 CLPLTC 267

## RESULT 12

US-09-983-025-25  
; Sequence 25, Application US/09983025  
; Publication No. US20030124529A1  
; GENERAL INFORMATION:  
; APPLICANT: OXVIG, Claus  
; APPLICANT: OVERGAARD, Michael T.  
; TITLE OF INVENTION: PREGNANCY-ASSOCIATED PLASMA PROTEIN-A2 (PAPP-A2)  
; FILE REFERENCE: OXVIG-1A  
; CURRENT APPLICATION NUMBER: US/09/983,025  
; CURRENT FILING DATE: 2001-10-22  
; PRIOR APPLICATION NUMBER: US 60/241,840  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: DK PA 2000 01571  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 1627  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-983-025-25

Query Match 92.1%; Score 35; DB 10; Length 1627;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVPLTC 6

Db 1410 CVPVTC 1415

## RESULT 13

US-10-295-027-663  
; Sequence 663, Application US/10295027  
; Publication No. US20030232350A1  
; GENERAL INFORMATION:  
; APPLICANT: Afar, Daniel  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsberg, Wendy M.  
; APPLICANT: Gish, Kurt C.  
; APPLICANT: Glynn, Richard  
; APPLICANT: Hevezi, Peter A.  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan R.  
; APPLICANT: Eos Biotechnology, Inc.  
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
; FILE REFERENCE: 018501-01250005  
; CURRENT APPLICATION NUMBER: US/10/295,027  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: US 09/663,733  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: US 60/350,666  
; PRIOR FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: US 60/335,394  
; PRIOR FILING DATE: 2001-11-15  
; PRIOR APPLICATION NUMBER: US 60/332,464  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 60/334,393  
; PRIOR FILING DATE: 2001-11-29

```
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 663
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-663

Query Match          92.1%; Score 35; DB 15; Length 1627;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVPLTC 6
Db      1410 CVPVTC 1415

RESULT 14
US-10-424-599-272858
; Sequence 272858, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 272858
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_88413C.1.pep
US-10-424-599-272858

Query Match          89.5%; Score 34; DB 12; Length 80;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVPLTC 6
Db      40 CVPVLC 45

RESULT 15
US-10-104-047-3842
; Sequence 3842, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3842
; LENGTH: 570
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3842

Query Match          89.5%; Score 34; DB 15; Length 570;
Best Local Similarity 83.3%; Pred. No. 6.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 CVPLTC 6
Db      186 CVPVLC 191

Search completed: September 5, 2004, 10:29:22
Job time : 17.8485 secs
```

**This Page Blank (uspto)**

**..... Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:55:30 ; Search time 5.33333 Seconds  
(without alignments)  
58.079 Million cell updates/sec

Title: US-09-761-636A-12

Perfect score: 38

Sequence: 1 CVPLTC 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/iaa/5A COMB.pdp.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pdp.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pdp.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pdp.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pdp.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	86.8	159	4	US-09-328-352-4708
2	33	86.8	549	1	US-08-325-071-61
3	33	86.8	549	3	US-08-461-004A-61
4	33	86.8	620	1	US-08-325-071-65
5	33	86.8	620	3	US-08-461-004A-65
6	33	86.8	650	1	US-08-325-071-56
7	33	86.8	650	1	US-08-325-071-59
8	33	86.8	650	1	US-08-325-071-63
9	33	86.8	650	1	US-08-325-071-67
10	33	86.8	650	3	US-08-461-004A-56
11	33	86.8	650	3	US-08-461-004A-59
12	33	86.8	650	3	US-08-461-004A-63
13	33	86.8	650	3	US-08-461-004A-67
14	33	86.8	688	1	US-08-325-071-57
15	33	86.8	688	3	US-08-461-004A-57
16	33	86.8	1587	4	US-09-845-583A-10
17	33	86.8	1587	4	US-09-561-709B-3
18	32	84.2	38	6	5208144-23
19	32	84.2	49	4	US-09-904-615-83
20	32	84.2	726	6	5208144-37
21	32	84.2	1761	4	US-09-561-709B-1
22	32	84.2	3571	4	US-09-911-842A-2
23	32	84.2	4654	3	US-08-476-515A-84
24	32	84.2	4655	3	US-08-652-877-84
25	32	84.2	4655	3	US-08-652-877-86
26	32	84.2	4655	3	US-08-652-877-88
27	32	84.2	4655	3	US-08-652-877-90

28 31 81.6 39 1 US-08-212-236-9 Sequence 9, Appli  
29 31 81.6 40 1 US-08-033-873-5 Sequence 5, Appli  
30 31 81.6 40 2 US-08-356-832-5 Sequence 5, Appli  
31 31 81.6 40 3 US-08-988-705-5 Sequence 5, Appli  
32 31 81.6 274 4 US-09-252-991A-17751 Sequence 17751, A  
33 31 81.6 321 4 US-09-252-991A-32545 Sequence 32545, A  
34 31 81.6 431 4 US-09-252-991A-18787 Sequence 18787, A  
35 31 81.6 810 2 US-08-820-170A-34 Sequence 34, Appl  
36 31 81.6 810 3 US-09-055-699-34 Sequence 34, Appl  
37 31 81.6 810 3 US-09-273-565-34 Sequence 34, Appl  
38 31 81.6 810 4 US-09-565-538-34 Sequence 34, Appl  
39 31 81.6 810 4 US-09-661-468-34 Sequence 34, Appl  
40 31 81.6 810 4 US-09-976-165-34 Sequence 34, Appl  
41 31 81.6 873 4 US-09-540-824-28 Sequence 28, Appl  
42 30 78.9 38 1 US-07-603-451A-1 Sequence 1, Appli  
43 30 78.9 38 1 US-08-060-822A-1 Sequence 1, Appli  
44 30 78.9 38 1 US-08-033-873-8 Sequence 8, Appli  
45 30 78.9 38 1 US-08-033-873-15 Sequence 15, Appli

## ALIGNMENTS

RESULT 1  
US-09-328-352-4708  
; Sequence 4708, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: GTC99-03PA  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4708  
; LENGTH: 159  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4708

Query Match 86.8%; Score 33; DB 4; Length 159;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
DB 113 CVPLAC 118

RESULT 2  
US-08-325-071-61  
; Sequence 61, Application US/08325071  
; Patent No. 5587311  
; GENERAL INFORMATION:  
; APPLICANT: COBON, Stewart Gary  
; APPLICANT: MOORE, Joanna Terry  
; APPLICANT: JOHNSON, Law Anthony York  
; APPLICANT: WILLADSEN, Peter  
; APPLICANT: KEMP, David Harold  
; APPLICANT: SRISKANTHA, Alagacone  
; APPLICANT: RIDING, George Alfred  
; APPLICANT: RAND, Keith No. 5587311man  
; TITLE OF INVENTION: DNA Encoding A Cell Membrane  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PI4912
FILING DATE: 16-OCT-1987
APPLICATION NUMBER: AU PI2570
FILING DATE: 19-JUN-1987
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-61

Query Match 86.8%; Score 33; DB 1; Length 549;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 66 CVPTTC 71

RESULT 3
US-08-461-004A-61
; Sequence 61, Application US/08461004A
; Patent No. 6235283
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 6235283man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,071
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/062,109
FILING DATE: 17-MAY-1993
APPLICATION NUMBER: US 07/926,368
FILING DATE: 07-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/242,196
FILING DATE: 06-JUL-1988
APPLICATION NUMBER: PCT/AU87/00401
FILING DATE: 27-NOV-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PI4912
FILING DATE: 16-OCT-1987
APPLICATION NUMBER: AU PI2570
FILING DATE: 19-JUN-1987
APPLICATION NUMBER: AU PH9196
FILING DATE: 27-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 60042/111 BIAU
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 61:
SEQUENCE CHARACTERISTICS:
LENGTH: 549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-071-61

Query Match 86.8%; Score 33; DB 3; Length 549;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 66 CVPTTC 71

RESULT 4
US-08-325-071-65
; Sequence 65, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.

;; TITLE OF INVENTION: Glycoprotein Of A Tick Gut  
;; NUMBER OF SEQUENCES: 71  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 3000 K Street, N.W.  
;; CITY: Washington, D.C.  
;; COUNTRY: USA  
;; ZIP: 20007-5109  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/325,071  
;; FILING DATE: 14-OCT-1993  
;; PRIOR APPLICATION DATA: US 08/062,109  
;; APPLICATION NUMBER: US 07/926,368  
;; FILING DATE: 17-MAY-1993  
;; PRIOR APPLICATION DATA: US 07/926,368  
;; APPLICATION NUMBER: 07/242,196  
;; FILING DATE: 06-JUL-1988  
;; APPLICATION DATA: PCT/AU87/00401  
;; FILING DATE: 27-NOV-1987  
;; APPLICATION NUMBER: AU P14912  
;; FILING DATE: 16-OCT-1987  
;; PRIOR APPLICATION DATA: AU P12570  
;; FILING DATE: 19-JUN-1987  
;; APPLICATION NUMBER: AU P12570  
;; FILING DATE: 27-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 60042/111 BIAU  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202 672 5300  
;; TELEFAX: 202 672 5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 65:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 620 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-325-071-65

Query Match 86.8%; Score 33; DB 1; Length 620;  
Best Local Similarity 83.3%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
Db 117 CVPTTC 122

RESULT 5  
US-08-461-004A-65  
; Sequence 65, Application US/08461004A  
; Patent No. 6235283  
; GENERAL INFORMATION:  
; APPLICANT: COBON, Stewart Gary  
; APPLICANT: MOORE, Joanna Terry  
; APPLICANT: JOHNSON, Law Anthony York  
; APPLICANT: WILLADSEN, Peter  
; APPLICANT: KEMP, David Harold  
; APPLICANT: SRISKANTHA, Alagacone

;; APPLICANT: RIDING, George Alfred  
;; APPLICANT: RAND, Keith No. 6235283man  
;; TITLE OF INVENTION: DNA Encoding A Cell Membrane  
;; TITLE OF INVENTION: Glycoprotein Of A Tick Gut  
;; NUMBER OF SEQUENCES: 71  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Foley & Lardner  
;; STREET: 3000 K Street, N.W.  
;; CITY: Washington, D.C.  
;; COUNTRY: USA  
;; ZIP: 20007-5109  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/461,004A  
;; FILING DATE: 04-JUN-1995  
;; PRIOR APPLICATION DATA: US 08/325,071  
;; APPLICATION NUMBER: 19-OCT-1994  
;; FILING DATE: 17-MAY-1993  
;; APPLICATION NUMBER: US 08/062,109  
;; PRIOR APPLICATION DATA: US 07/926,368  
;; FILING DATE: 07-AUG-1992  
;; APPLICATION NUMBER: 07/242,196  
;; FILING DATE: 06-JUL-1988  
;; APPLICATION DATA: PCT/AU87/00401  
;; FILING DATE: 27-NOV-1987  
;; APPLICATION NUMBER: AU P14912  
;; FILING DATE: 16-OCT-1987  
;; PRIOR APPLICATION DATA: AU P12570  
;; FILING DATE: 19-JUN-1987  
;; APPLICATION NUMBER: AU P12570  
;; FILING DATE: 27-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 60042/152  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202 672 5300  
;; TELEFAX: 202 672 5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 65:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 620 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-461-004A-65

Query Match 86.8%; Score 33; DB 3; Length 620;  
Best Local Similarity 83.3%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6  
Db 117 CVPTTC 122

RESULT 6  
US-08-325-071-56  
; Sequence 56, Application US/08325071  
; Patent No. 5587311  
; GENERAL INFORMATION:

```

; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P14912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P12570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-071-56

Query Match 86.8%; Score 33; DB 1; Length 650;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6
DB 147 CVPTTC 152

RESULT 7
US-08-325-071-59

```



```
RESULT 8
US-08-325-071-63
; Sequence 63, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; APPLICATION NUMBER: AU P14912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P12570
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-071-63
Query Match 86.8%; Score 33; DB 1; Length 650;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CVPLTC 6

RESULT 9
US-08-325-071-67
; Sequence 67, Application US/08325071
; Patent No. 5587311
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 5587311man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,071
; FILING DATE: 14-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,109
; FILING DATE: 17-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; APPLICATION NUMBER: AU P14912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P12570
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/111 BIAU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-071-67
Query Match 86.8%; Score 33; DB 1; Length 650;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CVPLTC 6
```

```
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CVPLTC 6
Db 147 CVPTTC 152

Query Match 86.8%; Score 33; DB 3; Length 650;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVPLTC 6
Db 147 CVPTTC 152

RESULT 11
US-08-461-004A-56
; Sequence 56, Application US/08461004A
; Patent No. 6235283
; GENERAL INFORMATION:
; APPLICANT: COBON, Stewart Gary
; APPLICANT: MOORE, Joanna Terry
; APPLICANT: JOHNSON, Law Anthony York
; APPLICANT: WILLADSEN, Peter
; APPLICANT: KEMP, David Harold
; APPLICANT: SRISKANTHA, Alagacone
; APPLICANT: RIDING, George Alfred
; APPLICANT: RAND, Keith No. 6235283man
; TITLE OF INVENTION: DNA Encoding A Cell Membrane
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,004A
; FILING DATE: 04-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/325,071
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,368
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/242,196
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P12570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU P14912
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/152
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
```



;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU P12570  
;; FILING DATE: 19-JUN-1987  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU PH9196  
;; FILING DATE: 27-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 60042/152  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202 672 5300  
;; TELEFAX: 202 672 5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 67:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 650 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-461-004A-67

Query Match 86.8%; Score 33; DB 3; Length 650;  
Best Local Similarity 83.3%; Pred. No. 4.4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTC 6  
||| ||  
Db 147 CVPTTC 152

## RESULT 14

US-08-325-071-57  
; Sequence 57, Application US/08325071  
; Patent No. 5587311  
; GENERAL INFORMATION:  
; APPLICANT: COBON, Stewart Gary  
; APPLICANT: MOORE, Joanna Terry  
; APPLICANT: JOHNSON, Law Anthony York  
; APPLICANT: WILLADSEN, Peter  
; APPLICANT: KEMP, David Harold  
; APPLICANT: SRISKANTHA, Alagacone  
; APPLICANT: RIDING, George Alfred  
; APPLICANT: RAND, Keith No. 5587311man  
; TITLE OF INVENTION: DNA Encoding A Cell Membrane  
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325,071  
; FILING DATE: 14-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/062,109  
; FILING DATE: 17-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,368  
; FILING DATE: 07-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/242,196  
; FILING DATE: 06-JUL-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU87/00401  
; FILING DATE: 27-NOV-1987

;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU P14912  
;; FILING DATE: 16-OCT-1987  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU P12570  
;; FILING DATE: 19-JUN-1987  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: AU PH9196  
;; FILING DATE: 27-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 60042/111 BIAU  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202 672 5300  
;; TELEFAX: 202 672 5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 57:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 688 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; US-08-325-071-57

Query Match 86.8%; Score 33; DB 1; Length 688;  
Best Local Similarity 83.3%; Pred. No. 4.6e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVPLTC 6  
||| ||  
Db 157 CVPTTC 162

## RESULT 15

US-08-461-004A-57  
; Sequence 57, Application US/08461004A  
; Patent No. 6235283  
; GENERAL INFORMATION:  
; APPLICANT: COBON, Stewart Gary  
; APPLICANT: MOORE, Joanna Terry  
; APPLICANT: JOHNSON, Law Anthony York  
; APPLICANT: WILLADSEN, Peter  
; APPLICANT: KEMP, David Harold  
; APPLICANT: SRISKANTHA, Alagacone  
; APPLICANT: RIDING, George Alfred  
; APPLICANT: RAND, Keith No. 6235283man  
; TITLE OF INVENTION: DNA Encoding A Cell Membrane  
; TITLE OF INVENTION: Glycoprotein Of A Tick Gut  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/461,004A  
; FILING DATE: 04-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/325,071  
; FILING DATE: 19-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/062,109  
; FILING DATE: 17-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/926,368  
; FILING DATE: 07-AUG-1992  
; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER: 07/242,196
; FILING DATE: 06-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU87/00401
; FILING DATE: 27-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PI4912
; FILING DATE: 16-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PI2570
; FILING DATE: 19-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PH9196
; FILING DATE: 27-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 60042/152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-461-004A-57

```

```

Query Match      86.8%; Score 33; DB 3; Length 688;
Best Local Similarity 83.3%; Pred No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 CVPLTC 6
Db      157 CVPTTC 162

```

```

Search completed: September 5, 2004, 10:22:00
Job time : 6.33333 secs

```

**This Page Blank (uspto)**

**Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:47:29 ; Search time 7.67677 Seconds  
(without alignments)  
125.302 Million cell updates/sec

Title: US-09-761-636A-13  
Perfect score: 56  
Sequence: 1 CISVPLSVPC 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: piri:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	69.6	85	T44555	hypothetical prote
2	39	69.6	148	G82599	hypothetical prote
3	38	67.9	502	F86298	IMP dehydrogenase
4	38	67.9	503	JC4999	hypothetical prote
5	38	67.9	616	D71864	hypothetical prote
6	37	66.1	89	F82521	hypothetical prote
7	37	66.1	485	A33647	sulfated surface g
8	36	64.3	134	D45835	Ly6 homolog RK3 pr
9	36	64.3	207	B86781	hypothetical prote
10	36	64.3	256	G75515	conserved hypothet
11	36	64.3	496	G83136	probable amino aci
12	36	64.3	512	H96759	probable steroid 2
13	36	64.3	581	S46742	hypothetical prote
14	36	64.3	669	T08827	hypothetical prote
15	35	62.5	100	F72751	hypothetical prote
16	35	62.5	255	JC7593	SH2 domain-contain
17	35	62.5	323	I49529	transcription fact
18	35	62.5	372	S72711	masC protein - Myc
19	35	62.5	382	A83171	probable aminotran
20	35	62.5	438	T15039	omega-3 fatty acid
21	35	62.5	468	T50982	origin recognition
22	35	62.5	503	D75437	DNA repair protein
23	35	62.5	578	A86926	probable acyl-coA
24	35	62.5	579	B86926	acyl-CoA synthetas
25	35	62.5	579	B86926	probable acyl-CoA
26	35	62.5	580	B70668	probable single-st
27	35	62.5	706	D71289	hypothetical prote
28	35	62.5	759	G86506	hypothetical prote
29	35	62.5	759	G72115	hypothetical prote

30 35 62.5 1162 2 B97852  
31 34 60.7 38 2 C54471  
32 34 60.7 77 2 D97718  
33 34 60.7 156 2 H71023  
34 34 60.7 249 2 AD0966  
35 34 60.7 250 2 S06314  
36 34 60.7 252 2 B33538  
37 34 60.7 328 2 F89914  
38 34 60.7 387 2 T44873  
39 34 60.7 399 2 AD2559  
40 34 60.7 410 2 I38502  
41 34 60.7 440 2 I48291  
42 34 60.7 461 1 KXHU  
43 34 60.7 502 2 C95349  
44 34 60.7 527 2 H85135  
45 34 60.7 554 2 JE0303

hypothetical prote  
agitoxin 3 - scorp  
acetate kinase (Ac  
hypothetical prote  
probable PTS syste  
regulatory protein  
hypothetical prote  
probable secreted  
hypothetical prote  
gene Brn-3b protei  
transcription fact  
protein C (activat  
hypothetical prote  
hypothetical prote  
propanediol dehydr

ALIGNMENTS

RESULT 1  
T44555  
hypothetical protein PA0631 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 06-Oct-2000  
C/Accession: T44555; E83568  
R;Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.; Oht  
submitted to the EMBL Data Library, August 1999  
A:Description: Genetic relationship between bacteriocins and bacteriophages.  
A:Reference number: Z22790  
A:Accession: T44555  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-85 <NAK>  
A:Cross-references: EMBL:AB030825; PIDN:BAA83170.1  
A:Experimental source: strain PA01  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: E83568  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-85 <STO>  
A:Cross-references: GB:AE004498; GB:AE004091; PIDN:G9946491; PIDN:AG04020.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0631

Query Match 69.6%; Score 39; DB 2; Length 85;  
Best Local Similarity 66.7%; Pred. No. 6;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISVPLSVPC 10  
: |||: |||  
DB 27 VEVLAVPC 35

RESULT 2

G82599  
hypothetical protein XF2118 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C/Accession: G82599  
R;Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A>Note: for a complete list of authors see reference number A59328 below  
A:Accession: G82599

A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <SIM>  
A:Cross-references: GB:AE004026; NID:g9107228; PIDN:AAF84917.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XP2118

Query Match 69.6%; Score 39; DB 2; Length 148;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
|:|:|:|:  
Db 8 CCAVPMNAPC 17

RESULT 3  
F86298  
IMP dehydrogenase (EC 1.1.1.205) - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C;Accession: F86298  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: F86298  
A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-502 <STO>  
A:Cross-references: GB:AE005172; NID:g4966356; PIDN:AAD34687.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1  
C;Superfamily: IMP dehydrogenase; CBS homology; IMP dehydrogenase amino-terminal homolog  
C;Keywords: GMP biosynthesis; NAD; oxidoreductase

Query Match 67.9%; Score 38; DB 2; Length 502;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPLSVPC 10  
|:|:|:|:  
Db 49 VPLSIPC 55

RESULT 4  
JC4999  
IMP dehydrogenase (EC 1.1.1.205) - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: JC4999

A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <SIM>  
A:Cross-references: GB:AE003849; NID:g9107228; PIDN:AAF84917.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XP2118

Query Match 69.6%; Score 39; DB 2; Length 148;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
|:|:|:|:  
Db 8 CCAVPMNAPC 17

RESULT 3  
F86298  
IMP dehydrogenase (EC 1.1.1.205) - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C;Accession: F86298  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: F86298  
A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-502 <STO>  
A:Cross-references: GB:AE005172; NID:g4966356; PIDN:AAD34687.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1  
C;Superfamily: IMP dehydrogenase; CBS homology; IMP dehydrogenase amino-terminal homolog  
C;Keywords: GMP biosynthesis; NAD; oxidoreductase

Query Match 67.9%; Score 38; DB 2; Length 502;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPLSVPC 10  
|:|:|:|:  
Db 49 VPLSIPC 55

RESULT 4  
JC4999  
IMP dehydrogenase (EC 1.1.1.205) - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: JC4999

A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <SIM>  
A:Cross-references: GB:AE004026; NID:g9107228; PIDN:AAF84917.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XP2118

Query Match 69.6%; Score 39; DB 2; Length 148;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
|:|:~:~:~:~:  
Db 8 CCAVPMNAPC 17

RESULT 3  
F86298  
IMP dehydrogenase (EC 1.1.1.205) - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002  
C;Accession: F86298  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A



A:Molecule type: DNA  
 A:Residues: 1-89 <SIM>  
 A:Cross-references: GB:AE004080; GB:AE003849; NID:g9107971; PIDN:AA085521.1; GSPDB:GN001  
 A:Experimental source: strain 945C  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Canargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF2736

Query Match 66.1%; Score 37; DB 2; Length 89;  
 Best Local Similarity 66.7%; Pred. No. 14;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ISVPLSVPC 10  
 : |||||  
 Db 19 VSAPLRVPC 27

RESULT 7  
 A33647  
 sulfated surface glycoprotein 185 - Volvox carteri  
 C:Species: Volvox carteri  
 C:Date: 11-Apr-1990 #sequence\_revision 11-Apr-1990 #text\_change 21-Jul-2000  
 C:Accession: A33647  
 R:Ertl, H.; Mengelle, R.; Wenzl, S.; Engel, J.; Sumper, M.  
 J. Cell Biol. 109, 3493-3501, 1989  
 A:Title: The extracellular matrix of Volvox carteri: molecular structure of the cellular  
 A:Reference number: A33647; MUID:90094551; PMID:2689458  
 A:Accession: A33647  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-485 <ERT>  
 A:Cross-references: GB:X51616; NID:g21999; PIDN:CAA35953.1; PID:g1405821  
 C:Keywords: glycoprotein

Query Match 66.1%; Score 37; DB 2; Length 485;  
 Best Local Similarity 60.8%; Pred. No. 73;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 10  
 : |||||  
 Db 83 CQTVPLTEPC 92

RESULT 8  
 D45835  
 Ly6 homolog RK3 precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 31-Jan-2000  
 C:Accession: D45835; C45835  
 R:Friedman, S.; Palfrey, R.G.E.; Sirlin, S.; Haemmerling, U.  
 Immunogenetics 31, 104-111, 1990  
 A:Title: Analysis of three distinct Ly6-A-related cDNA sequences isolated from rat kidney  
 A:Reference number: A45835; MUID:90152758; PMID:2154400  
 A:Accession: D45835  
 A:Molecule type: mRNA  
 A:Residues: 1-134 <FRI>  
 A:Cross-references: GB:M30690; NID:g205249; PIDN:AAA41547.1; PID:g205250  
 A:Experimental source: clone RK3  
 A:Accession: C45835  
 A:Molecule type: mRNA

A:Residues: 2-134 <FR2>  
 A:Cross-references: GB:M30691; NID:g205251; PIDN:AAA41548.1; PID:g205252  
 A:Experimental source: Clone RK11  
 C:Superfamily: Ly-6 antigen; Ly-6 homology  
 C:Keywords: blocked carboxyl end; glycoprotein; lipoprotein; phosphatidylinositol linka  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-105/Product: Ly6 homolog RK3 #status predicted <MAT>  
 F:106-134/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
 F:105/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature for

Query Match 64.3%; Score 36; DB 2; Length 134;  
 Best Local Similarity 60.0%; Pred. No. 31;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 10  
 : |||||  
 Db 32 CIEVPLNANC 41

RESULT 9  
 B86781  
 hypothetical protein ymhC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: B86781  
 R:Bolotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: B86781  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-207 <STO>  
 A:Cross-references: GB:AE005176; PID:g12724223; PIDN:AAK05348.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: ymhC

Query Match 64.3%; Score 36; DB 2; Length 207;  
 Best Local Similarity 55.6%; Pred. No. 48;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 9  
 : |||||  
 Db 128 CINLPLOIP 136

RESULT 10  
 G75515  
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: G75515  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: G75515  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-256 <WHI>  
 A:Cross-references: GB:AE001906; NID:g6458151; PIDN:AAF10050.1; PID:g645815  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR0471  
 A:Map position: 1

Query Match 64.3%; Score 36; DB 2; Length 256;  
 Best Local Similarity 50.0%; Pred. No. 59;  
 Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
 Db 4 CVPAPRSAPC 13

RESULT 11

G83136  
 probable amino acid permease PA4072 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: G83136  
 R:Scover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; PMID:10984043  
 A:Accession: G83136  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-496 <STO>  
 A:Cross-references: GB:AE004823; GB:AE004091; NID:9950265; PIDN:AAG07459.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4072

Query Match 64.3%; Score 36; DB 2; Length 496;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
 Db 345 C1AVAAIIPC 354

RESULT 12

H96759  
 probable steroid 22-alpha-hydroxylase T9L24.44 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Apr-2002  
 C:Accession: H96759  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huiziar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 A:Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; PMID:121016719; PMID:11130712  
 A:Accession: H96759  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-512 <STO>  
 A:Cross-references: GB:AE005173; NID:g1120803; PIDN:AAG30983.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: T9L24.44  
 A:Map position: 1  
 A:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology  
 C:Keywords: heme; iron; metalloprotein  
 F:455/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 64.3%; Score 36; DB 2; Length 512;  
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 9  
 Db 244 CLSVPIDLP 252

RESULT 13

S46742  
 hypothetical protein YHR032w - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein H8179.15  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Apr-2002  
 C:Accession: S46742  
 R:Du, Z.  
 submitted to the EMBL Data Library, May 1994  
 A:Description: The sequence of S. cerevisiae cosmid 8179.  
 A:Reference number: S46732  
 A:Accession: S46742  
 A:Molecule type: DNA  
 A:Residues: 1-581 <DUZ>  
 A:Cross-references: EMBL:U00062; NID:g488162; PIDN:AAB68911.1; PID:g488176; MIPS:YHR032w  
 C:Genetics:  
 A:Cross-references: SGD:S0001074  
 A:Map position: 8R  
 C:Keywords: transmembrane protein

Query Match 64.3%; Score 36; DB 2; Length 581;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
 Db 564 CVPSPVSSGC 573

RESULT 14

T08827  
 hypothetical protein cotel - human  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: T08827  
 R:Winfield, S.H.; Tayebi, N.; Martin, B.M.; Ginns, E.I.; Sidransky, E.  
 Genome Res. 7, 1020-1026, 1997  
 A:Title: Identification of three additional genes contiguous to the glucocerebrosidase  
 A:Reference number: 216482; PMID:9747496; PMID:9331372  
 A:Accession: T08827  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-669 <WIN>  
 A:Cross-references: EMBL:AF023268; NID:g2564910; PIDN:AAC51822.1; PID:g2564916  
 C:Genetics:  
 A:Gene: cotel  
 A:Map position: 1  
 A:Introns: 75/3; 94/3; 131/3; 171/3; 207/3; 266/2; 299/3; 323/1; 505/1; 528/1; 612/3

Query Match 64.3%; Score 36; DB 2; Length 669;  
 Best Local Similarity 70.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
 Db 139 CVPSPVLRPC 148

RESULT 15

F72751  
 hypothetical protein APE0003 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: F72751  
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tak  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy  
 A:Reference number: A72450; PMID:99310339; PMID:10382966  
 A:Accession: F72751  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-100 <KAW>

A:Cross-references: DDBJ:AP000058; NID:G5103388; PIDN:EAA78912.1; PID:G5103391  
A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0003

C:Superfamily: Aeropyrum pernix hypothetical protein APE0003

Query Match 62.5%; Score 35; DB 2; Length 100;  
Best Local Similarity 71.4%; Pred. No. 35;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VPLSVPC 10

Db 57 IPLTVPC 63

Search completed: September 5, 2004, 10:01:24  
Job time : 9.67677 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:38:39 ; Search time 4.54545 Seconds  
(without alignments)  
114.554 Million cell updates/sec

Title: US-09-761-636A-13  
Perfect score: 56  
Sequence: 1 C1SVPLSVPC 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	42	75.0	569	1 MANC_PIRSP	P55298 piromyces s
2	42	75.0	571	1 MANB_PIRSP	P55297 piromyces s
3	38	67.9	502	1 IMH2 ARATH	Q98a34 arabidopsis
4	38	67.9	503	1 IMH1 ARATH	P47996 arabidopsis
5	38	67.9	616	1 Y396 HELPJ	Q9zkt3 helicobacte
6	37	66.1	420	1 CG2A CHLVR	P51986 chlorohydra
7	37	66.1	485	1 SSGP VOLCA	F21997 volvox cart
8	36	64.3	581	1 YHJ2 YEAST	F38767 saccharomyc
9	36	64.3	669	1 COTE HUMAN	P81408 homo sapien
10	35	62.5	298	1 YMA2 MYCBO	Q02278 mycobacteri
11	35	62.5	323	1 MAFB HUMAN	Q9y5q3 homo sapien
12	35	62.5	323	1 MAFB MOUSE	P54841 mus musculu
13	35	62.5	323	1 MAFB RAT	P54842 rattus norv
14	35	62.5	369	1 V181 FOWPV	Q9j552 fowlpox vir
15	35	62.5	578	1 FD21 MYCTU	O50441 mycobacteri
16	35	62.5	579	1 FD21 MYCLE	P54200 mycobacteri
17	35	62.5	1158	1 KCH2 CANFA	Q9tsz3 canis fami
18	34.5	61.6	120	1 SY23 HUMAN	P55773 homo sapien
19	34	60.7	38	1 SC33 LEIQH	P46112 leleirus qui
20	34	60.7	250	1 LUXR VIBFI	P12746 vibrio fiesc
21	34	60.7	336	1 V027 FOWPV	Q9j5h4 fowlpox vir
22	34	60.7	410	1 PO42 HUMAN	Q12837 homo sapien
23	34	60.7	433	1 RTCI DICDI	O15746 dictyosteli
24	34	60.7	461	1 PRTC HUMAN	P04070 homo sapien
25	34	60.7	554	1 DHAB SALTY	P37450 salmonella
26	34	60.7	590	1 MP44 SFVKA	Q9q926 shope fibro
27	34	60.7	595	1 SILL HUMAN	Q96pg1 homo sapien
28	34	60.7	597	1 SILL PANTR	Q95lho pan troglod
29	34	60.7	606	1 MANA_PIRSP	P55296 piromyces s
30	34	60.7	829	1 CADG HUMAN	O75309 homo sapien
31	33	58.9	348	1 OPSB_HUMAN	P03999 homo sapien
32	33	58.9	348	1 OPSB_PANTR	P60015 pan troglod
33	33	58.9	349	1 OPSB_SAIBB	O13092 saimiri bol

34	33	58.9	369	1 MAF RAT	P54844 rattus norv
35	33	58.9	369	1 TMAF_AVIS4	P23091 avian muscu
36	33	58.9	370	1 MAF_MOUSE	P54843 mus musculu
37	33	58.9	403	1 MAF_HUMAN	O75444 homo sapien
38	33	58.9	422	1 Y140 HUMAN	Q14153 homo sapien
39	33	58.9	465	1 YGT4_YEAST	P53101 saccharomyc
40	33	58.9	549	1 WD23_MOUSE	Q91vu6 mus musculu
41	33	58.9	573	1 C114_MOUSE	P19467 mus musculu
42	33	58.9	583	1 FD25 MYCTU	Q50586 mycobacteri
43	33	58.9	662	1 T9S2_MOUSE	P58021 mus musculu
44	33	58.9	663	1 T9S2_HUMAN	Q95805 homo sapien
45	32	57.1	61	1 IBPI_TACTR	P16044 tachypleus

## ALIGNMENTS

RESULT 1  
MANC\_PIRSP STANDARD; PRT; 569 AA.  
AC P55298;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Mannan endo-1,4-beta-mannosidase C precursor (EC 3.2.1.78) (Beta-mannanase C) (1,4-beta-D-mannan mannanohydrolase C).  
DE mannanase C (1,4-beta-D-mannan mannanohydrolase C).  
GN MANC.  
OS Piromyces sp.  
OC Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;  
OC Neocallimastacaceae; Piromyces.  
OX NCBI\_TaxID=45796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96333314; PubMed=8768520;  
RA Millward-Sadler S.J., Hall J., Black G.W., Hazlewood G.P., Gilbert H.J.;  
RT "Evidence that the Piromyces gene family encoding endo-1,4-mannanases arose through gene duplication.";  
RT FEMS Microbiol. Lett. 141:183-188(1996).  
RL -I- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans.  
CC -I- SIMILARITY: THE 39 AA REPEATS ARE SIMILAR TO THOSE IN MANA, MANC AND XYNA; AND TO THOSE OF N.PATRICIARUM XYNA.  
CC -I- SIMILARITY: Belongs to family 26 of glycosyl hydrolases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X97520; CAA66134.1; --  
CC InterPro: IPR002883; Dockerin CBD 5.  
CC InterPro: IPR009034; Dockerin CBD.  
CC InterPro: IPR008979; Gal bind like.  
CC InterPro: IPR008005; Glyco\_hydro\_26.  
CC Pfam: PF02013; CBM 10; 2.  
CC Pfam: PF02156; Glyco\_hydro\_26; 1.  
CC PRINTS: PR00739; GLHYDLASE26.  
CC HydroLase; Glycosidase; signal; MULTIGENE family; Repeat.  
KW SIGNAL  
FT CHAIN 1 18 POTENTIAL.  
FT CHAIN 19 569 MANNA ENDO-1,4-BETA-MANNOSIDASE C.  
FT DOMAIN 490 569 2 X 39 AA APPROXIMATE REPEATS.  
FT REPEAT 488 525 1.  
FT REPEAT 531 569 2.  
FT DOMAIN 473 477 POLY-ASN.  
FT DOMAIN 480 486 POLY-ASN.  
SQ SEQUENCE 569 AA; 64115 MW; 19277764E18328B5 CRC64;  
Query Match 75.0%; Score 42; DB 1; Length 569;

Best Local Similarity 60.0%; Pred. No. 4.3;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10

Db 490 CFSIPLGYPC 499

# RESULT 2

MANB\_PIRSP STANDARD; PRT; 571 AA.  
AC P5297;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Mannan endo-1,4-beta-mannosidase B precursor (EC 3.2.1.78) (Beta-mannanase B) (1,4-beta-D-mannan mannanohydrolase B).  
DE MANB.  
GN PIRMYCES sp.  
OS Eukaryota; Fungi; Chytridiomycota; Neocallimasticales;  
OC Neocallimastacaceae; Piriomyces.  
OC NCBI\_TaxID=45796;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=96313314; PubMed=8768520;  
RX Millward-Sadler S.J., Hall J., Black G.W., Hazlewood G.P.,  
RA Gilbert H.J.;  
RT "Evidence that the Piriomyces gene family encoding endo-1,4-mannanases arose through gene duplication."  
RL FEMS Microbiol Lett. 141:183-188 (1996).  
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans, glucomannans, and galactoglucomannans.  
CC -!- SIMILARITY: THE 39 AA REPEATS ARE SIMILAR TO THOSE IN MANA, MANC AND XYNA; AND TO THOSE OF N.PATRICIARUM XYNA.  
CC -!- SIMILARITY: Belongs to family 26 of glycosyl hydrolases.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; X97408; CAA66061.1; -  
DR InterPro; IPR002883; Dockerin CBD 5.  
DR InterPro; IPR009034; Dockering CBD.  
DR InterPro; IPR008979; Gal\_bind-like.  
DR InterPro; IPR008005; Glyco\_hydro\_26.  
DR Pfam; PF02013; CBM10; 2.  
DR Pfam; PF02156; Glyco\_hydro\_26; 1.  
DR PRINTS; PR00739; GLHYDRLASE26.  
KW Hydrolase; Glycosidase; Signal; Multigene family; Repeat.  
FT SIGNAL 1 19 POTENTIAL  
FT CHAIN 20 571  
FT DOMAIN 490 571  
FT REPEAT 490 527 1.  
FT REPEAT 533 571 2.  
FT DOMAIN 475 479 POLY-ASN.  
FT DOMAIN 482 488 POLY-ASN.  
SQ SEQUENCE 571 AA; 64397 MW; B13E44581FAA9DAA CRC64;

Query Match 75.0%; Score 42; DB 1; Length 571;  
Best Local Similarity 60.0%; Pred. No. 4.4;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10

Db 492 CFSIPLGYPC 501

# RESULT 3

IMH2\_ARATH STANDARD; PRT; 502 AA.  
ID IMH2\_ARATH  
AC Q9SA34;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Probable inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205) (IMP dehydrogenase) (IMPDH) (IMPD).  
DE AT1G16350 OR F309.15. (Mouse-ear cross).  
GN Arabidopsis thaliana.  
OS Arabidopsi viridiplantae;  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Columbia;  
RC MEDLINE=21016719; PubMed=11130712;  
RX Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziali A., Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana";  
RL Nature 408:816-820 (2000).  
CC -!- CATALYTIC ACTIVITY: Inosine 5'-phosphate + NAD(+) + H(2)O = xanthosine 5'-phosphate + NADH.  
CC -!- PATHWAY: First reaction unique to GMP biosynthesis.  
CC -!- SIMILARITY: Belongs to the IMPDH/GMPR family.  
CC -!- SIMILARITY: Contains 1 CBS domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AC006341; AAD34687.1; -  
DR PIR; F86298; F86298.  
DR HSSP; F12268; 1B30.  
DR InterPro; IPR003009; FMN enzyme.  
DR InterPro; IPR005990; IMP dehyd.  
DR InterPro; IPR001093; IMPDH/GMPrtase.  
DR Pfam; PF00478; IMPDH; 1.  
DR TIGRfams; TIGR01302; IMP dehydrog; 1.  
DR PROSITE; PS00487; IMP\_DH\_GMP\_RED; 1.  
KW Oxidoreductase; NAD; GMP biosynthesis; Purine biosynthesis; Repeat;  
KW CBS domain.  
FT DOMAIN 167 219 CBS.  
FT BINDING 321 321 IMP (POTENTIAL).  
SQ SEQUENCE 502 AA; 54051 MW; FB87D84160818310 CRC64;

Query Match 67.9%; Score 38; DB 1; Length 502;  
Best Local Similarity 85.7%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPLSVPC 10

Db 49 VPLSVPC 55

```

KW CBS domain. 322 322 IMP (POTENTIAL).
FT BINDING 503 AA; 54194 MW; ADDAF9C3A697A9A CRC64;
SQ SEQUENCE 503 AA; 54194 MW; ADDAF9C3A697A9A CRC64;

Query Match 67.9%; Score 38; DB 1; Length 503;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VPLSVPC 10
Db 49 VPLSIPC 55

RESULT 5
Y396_HELPJ STANDARD; PRT; 616 AA.
AC Q92KF3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein JHP0985.
GN JHP0985.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.B., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -!- SIMILARITY: Belongs to the ubiD family.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

-----
CC EMBL; AS001527; AAD06568.1; -.
CC PIR; D71864; D71864.
CC InterPro; IPR002830; carboxylase.
CC Pfam; PF01977; UbiD; 1.
CC TIGRFAMs; TIGR00148; TIGR00148; 1.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 616 AA; 70989 MW; C219E1DCE4C8BDD5 CRC64;

Query Match 67.9%; Score 38; DB 1; Length 616;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CISVPLSVP 9
Db 253 CLSNPLSVP 261

RESULT 6
CG2A_CHLVR
ID CG2A_CHLVR STANDARD; PRT; 420 AA.
AC P51986;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE G2/mitotic-specific cyclin A (fragment).
OS Chlorohydra viridissima (Hydra).

```

CC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;  
CC Hydridae; Chlorohydra.  
OX NCBI\_TaxID=6082;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96338615; PubMed=8743953;  
RA Schreulen-Bichle I., Hoffmeister S., Herrmans-Borgmeyer I.,  
RA Schaller H.C.;  
RT "Presence and expression of G2 cyclins in the coelenterate hydra.";  
RL J. Cell Sci. 109:1063-1069(1996).  
CC -!- FUNCTION: Essential for the control of the cell cycle at the G2/M  
CC (mitosis) transition. Interacts with the CDC2 and CDK2 protein  
CC kinases to form MPP. G2/M cyclins accumulate steadily during G2  
CC and are abruptly destroyed at mitosis (By similarity).  
CC -!- SIMILARITY: Belongs to the cyclin family. Cyclin AB subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X51616; CAA35953.1; -.  
CC PIR; A33647; A33647.  
KW Signal; Glycoprotein; Hydroxylation.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 485 SULFATED SURFACE GLYCOPROTEIN 185.  
FT DOMAIN 228 340 PRO-RICH.  
FT DOMAIN 260 295 POLY-PRO.  
FT CARBOHYD 193 193 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 485 AA; 50436 MW; A52216400A031421 CRC64;  
Query Match 66.1%; Score 37; DB 1; Length 485;  
Best Local Similarity 60.0%; Pred. No. 29;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CISVPLSVPC 10  
DB 83 CQVPLTEPC 92  
RESULT 8  
YHJ2\_YEAST STANDARD; PRT; 581 AA.  
ID YHJ2\_YEAST  
AC P38767;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 64.2 kDa protein in SLT2-PUT2 intergenic region.  
GN YHR032W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,  
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,  
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,  
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,  
RA Vaudin M.;  
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
RT VIII.";  
RL Science 265:2077-2082(1994).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the multi antimicrobial extrusion (MATE)  
CC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U00062; AAB68911.1; -.  
CC PIR; S46742; S46742.  
DR DR

OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;  
OX NCBI\_TaxID=6082;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96338615; PubMed=8743953;  
RA Schreulen-Bichle I., Hoffmeister S., Herrmans-Borgmeyer I.,  
RA Schaller H.C.;  
RT "Presence and expression of G2 cyclins in the coelenterate hydra.";  
RL J. Cell Sci. 109:1063-1069(1996).  
CC -!- FUNCTION: Essential for the control of the cell cycle at the G2/M  
CC (mitosis) transition. Interacts with the CDC2 and CDK2 protein  
CC kinases to form MPP. G2/M cyclins accumulate steadily during G2  
CC and are abruptly destroyed at mitosis (By similarity).  
CC -!- SIMILARITY: Belongs to the cyclin family. Cyclin AB subfamily.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X90983; CAA62470.1; -.  
CC HSSP; P30274; LVIN.  
DR InterPro; IPR006670; Cyclin.  
DR InterPro; IPR004367; Cyclin\_Cterm.  
DR InterPro; IPR006671; Cyclin\_N.  
DR Pfam; PF00134; Cyclin; 1.  
DR Pfam; PF02984; Cyclin; 1.  
DR SMART; SM00385; CYCLIN; 2.  
DR PROSITE; PS00292; CYCLINS; 1.  
KW Cyclin; Cell cycle; Cell division; Mitosis.  
FT NON TER 1  
SQ SEQUENCE 420 AA; 47989 MW; 79225BF8D1B44788 CRC64;  
Query Match 66.1%; Score 37; DB 1; Length 420;  
Best Local Similarity 58.3%; Pred. No. 25;  
Matches 7; Conservative 2; Mismatches 1; Indels 2; Gaps 1;  
QY 1 CIS--VPLSVPC 10  
DB 408 CVSLVPLEMPC 419  
RESULT 7  
SSGP\_VOLCA STANDARD; PRT; 485 AA.  
ID SSGP\_VOLCA  
AC P21997;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Sulfated surface glycoprotein 185 precursor (SSG 185).  
OS Volvox carteri.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Volvocaceae; Volvox.  
OX NCBI\_TaxID=3067;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=f. Nagariensis / HK10;  
RX MEDLINE=90094551; PubMed=2689458;  
RA Ertl H., Mengede R., Wenzl S., Engel J., Sumper M.;  
RT "The extracellular matrix of Volvox carteri: molecular structure of  
RT the cellular compartment.";  
RL J. Cell Biol. 109:3493-3501(1989).  
CC -!- FUNCTION: The extracellular matrix (ECM) of volvox contains  
CC insoluble fibrous layers that surround individual cells at a  
CC distance to form contiguous cellular compartments. SSG 185 is the  
CC monomeric precursor of this substructure (C3Z structure).  
CC  
CC -!- SUBUNIT: Polymer.  
CC -!- PTM: Intersubunit cross-links are formed between saccharide chains  
CC rather than between polypeptide chains.  
DR DR



```

DR  GermOnline; 139349; -.
DR  SGD; S0001074; YHR032W.
DR  InterPro; IPR002528; Mate.
DR  Pfam; PF01554; Mate; 2.
DR  TIGRFAMs; TIGR00797; mate; 1.
KW  Hypothetical protein; Transmembrane; Transport.
FT  TRANSMEM 173 193 POTENTIAL.
FT  TRANSMEM 249 269 POTENTIAL.
FT  TRANSMEM 282 302 POTENTIAL.
FT  TRANSMEM 346 366 POTENTIAL.
FT  TRANSMEM 472 492 POTENTIAL.
SQ  SEQUENCE 581 AA; 64209 MW; B0592C480589319B CRC64;

Query Match          64.3%; Score 36; DB 1; Length 581;
Best Local Similarity 50.0%; Pred. No. 53;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY  1 CISVPLSVPC 10
   | | | | |
Db  564 CVSIPVSSGC 573

RESULT 9
COTE_HUMAN
ID  COTE_HUMAN STANDARD; PRT; 669 AA.
AC  P81408;
DT  30-MAY-2000 (Rel. 39, Created)
DI  30-MAY-2000 (Rel. 39, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  COTE1 protein.
GN  CLORF2 OR COTE1.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Hippocampus;
RX  MEDLINE=97474796; PubMed=9331372;
RA  Winfield S.L., Tavebi N., Martin B.M., Ginns E.I., Sidransky E.;
RT  "Identification of three additional genes contiguous to the
RT  glucocerebrosidase locus on chromosome 1q21: implications for Gaucher
RT  disease.";
RL  Genome Res. 7:1020-1026(1997).
CC  -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -|- TISSUE SPECIFICITY: Widely expressed.
CC  -|- SIMILARITY: TO HUMAN KIAA0574.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AF023268; AAC51822.1; -.
DR  PIR; T08827; T08827.
DR  Genew; HGNC:1233; Clorf2.
DR  GO; GO:0016021; C:integral to membrane; NAS.
KW  Transmembrane.
FT  TRANSMEM 34 54 POTENTIAL.
FT  TRANSMEM 67 87 POTENTIAL.
FT  TRANSMEM 91 111 POTENTIAL.
FT  TRANSMEM 174 194 POTENTIAL.
FT  DOMAIN 244 250 POLY-PRO.
FT  DOMAIN 635 639 POLY-SER.
SQ  SEQUENCE 669 AA; 71482 MW; E30360AC9A4571E6 CRC64;

Query Match          64.3%; Score 36; DB 1; Length 669;
Best Local Similarity 70.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY  1 CISVPLSVPC 10
   | | | | |
Db  139 CPSVPLLRPC 148

RESULT 10
YMA2_MYCBO
ID  YMA2_MYCBO STANDARD; PRT; 298 AA.
AC  Q02278;
DT  01-JUN-1994 (Rel. 29, Created)
DT  01-JUN-1994 (Rel. 29, Last sequence update)
DI  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hypothetical 32.3 kDa protein in mas 5' region.
OS  Mycobacterium bovis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BCG;
RX  MEDLINE=92406887; PubMed=1527058;
RA  Mathur M., Kolattukudy P.E.;
RT  "Molecular cloning and sequencing of the gene for mycoceroic acid
RT  synthase, a novel fatty acid elongating multifunctional enzyme, from
RT  Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
RL  J. Biol. Chem. 267:19388-19395(1992).
CC  -|- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV1521.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M95808; AAA25367.1; -.
DR  InterPro; IPR000873; AMP-bind.
DR  Pfam; PF00501; AMP-binding; 1.
KW  Hypothetical protein; Transmembrane.
SQ  SEQUENCE 298 AA; 32286 MW; 22232DP5DBC6F371 CRC64;

Query Match          62.5%; Score 35; DB 1; Length 298;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY  2 ISVPLSVP 9
   | | | | |
Db  63 IAVPLSVP 70

RESULT 11
MAFB_HUMAN
ID  MAFB_HUMAN STANDARD; PRT; 323 AA.
AC  Q9Y5G3; Q9H1F1;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DI  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Transcription factor MafB (V-maf musculoaponeurotic fibrosarcoma
DE  oncogene homolog B).
GN  MAFB OR KRML.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Bone marrow;
RX  MEDLINE=99375320; PubMed=10444328;
RA  Wang P.W., Eisenbart J.D., Cordes S.P., Barsh G.S., Stoffel M.,
RA  Le Beau M.M.;
RT  "Human KRML (MAFB): cDNA cloning, genomic structure, and evaluation as
RT  a candidate tumor suppressor gene in myeloid leukemias.";

```

```

RL Genomics 59:275-281(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Clark L.N., Clark S.Y., Clee C.M.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harte J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehaealho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Murrison T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.M., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.B., Whittaker P., Willey D.B., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Raba S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Plays a pivotal role in regulating lineage-specific hematopoiesis by repressing Ets1-mediated transcription of erythroid-specific genes in myeloid cells (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the bZIP family. Maf subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF134157; AAD30106.1; -
DR EMBL; AL035665; CAB75863.1; -
DR EMBL; BC028098; AAX28098.1; -
DR EMBL; BC036689; AAH36689.1; -

```

Genew; HGNC:6408; MAFB.

GO; GO:0005634; C:nucleus; TAS.

GO; GO:0007423; P:sensory organ development; TAS.

InterPro; IPR008917; Euk\_transcr\_DNA.

InterPro; IPR004827; TF\_ZIP.

InterPro; IPR004826; TF\_Maf.

Pfam; PFO3131; bZIP\_Maf; 1.

SMART; SM00338; BRLZ; 1.

PROSITE; PS0217; bZIP; 1.

PROSITE; PS00036; bZIP\_BASIC; FALSE NEG.

Transcription regulation; Repressor; DNA-binding; Nuclear protein.

DNA\_BIND 238 264 BASIC MOTIF.

DOMAIN 266 287 LEUCINE-ZIPPER.

DOMAIN 131 143 POLY-HIS.

DOMAIN 158 167 POLY-HIS.

CONFLICT 52 52 A -> V (IN REF. 1).

CONFLICT 241 241 Q -> H (IN REF. 1).

SEQUENCE 323 AA; 35792 MW; AOF3C09F8936CB16 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 323;

Best Local Similarity 75.0%; Pred. No. 43;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SVPLSVPC 10

Db 57 STPLSTPC 64

RESULT 12

MAFB MOUSE

ID MAFB MOUSE STANDARD; PRT; 323 AA.

AC P54841;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Transcription factor MafB (V-maf musculoaponeurotic fibrosarcoma oncogene homolog B) (transcription factor MAFB) (Segmentation protein KR) (Kreisel).

DE MAFB OR MAF1 OR KRML.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI TaxID=10090;

RP SEQUENCE FROM N.A.

RP MEDLINE=95094266; PubMed=8001130;

Cordes S.P., Barsh G.S.;

"The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor.";

Cell 79:1025-1034(1994).

[2]

RP SEQUENCE FROM N.A.

RP STRAIN=FVB/N; TISSUE=Liver;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Raba S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Whiting J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Plays a pivotal role in regulating lineage-specific hematopoiesis by repressing Ets1-mediated transcription of erythroid-specific genes in myeloid cells (By similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -!- SIMILARITY: Belongs to the bZIP family. Maf subfamily.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF134157; AAD30106.1; -

DR EMBL; AL035665; CAB75863.1; -

DR EMBL; BC028098; AAX28098.1; -

DR EMBL; BC036689; AAH36689.1; -

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Plays a pivotal role in regulating lineage-specific
CC hematopoiesis by repressing ETS1-mediated transcription of
CC erythroid-specific genes in myeloid cells.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.
CC -!- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND
CC IN THE CAUDAL HINDRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),
CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT
CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY
CC THROUGH R6.
CC -!- SIMILARITY: Belongs to the bZIP family. Maf subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L36435; AAA65689.1; -.
CC EMBL; BC038256; AAH38256.1; -.
CC PIR; I49529; I49529.
CC TRANSFAC; T01439; -.
CC MGD; MGI:104555; MafB.
CC DR GO; GO:0003677; F:DNA binding; IDA.
CC DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; IDA.
CC DR GO; GO:0007379; P:segment specification; IMP.
CC DR InterPro; IPR008917; Euk_transcr_DNA.
CC DR InterPro; IPR004827; TF_Maf.
CC DR Pfam; PF03131; bZIP_Maf; 1.
CC DR SMART; SM00338; BRLZ; 1.
CC DR PROSITE; PS0217; bZIP; 1.
CC DR PROSITE; PS00036; bZIP_BASIC; FALSE_NEG.
CC KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
CC FT DNA_BIND 238 264
CC FT DOMAIN 266 287
CC FT LEUCINE-ZIPPER.
CC FT DOMAIN 131 143
CC FT POLY-HIS.
CC FT DOMAIN 158 167
CC FT POLY-HIS.
CC FT MUTAGEN 248 248 N->S: LOSS OF TRANSCRIPTIONAL ACTIVITY.
CC SQ SEQUENCE 323 AA; 35809 MW; D77AE07ABD9C2AD2 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 323;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10
DB 57 STPLSTPC 64

RESULT 13
MAFB_RAT
ID MAFB_RAT STANDARD; PRT; 323 AA.
AC P54842;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription factor MafB (V-maf musculoaponeurotic fibrosarcoma
DE oncogene homolog B) (Transcription factor MAF1).
GN MAFB OR MAF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Wistar; TISSUE=Liver;
RX MEDLINE=97190228; PubMed=9038383;
RA Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,
RA Kuboki Y., Nishizawa M., Nishi S.;

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Plays a pivotal role in regulating lineage-specific
CC hematopoiesis by repressing ETS1-mediated transcription of
CC erythroid-specific genes in myeloid cells.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.
CC -!- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND
CC IN THE CAUDAL HINDRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),
CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT
CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY
CC THROUGH R6.
CC -!- SIMILARITY: Belongs to the bZIP family. Maf subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U56241; AAB50062.1; -.
CC InterPro; IPR008917; Euk_transcr_DNA.
CC DR InterPro; IPR004827; TF_EZIP.
CC DR Pfam; PF03131; bZIP_Maf; 1.
CC DR SMART; SM00338; BRLZ; 1.
CC DR PROSITE; PS0217; bZIP; 1.
CC DR PROSITE; PS00036; bZIP_BASIC; FALSE_NEG.
CC KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
CC FT DNA_BIND 238 264
CC FT DOMAIN 266 287
CC FT LEUCINE-ZIPPER.
CC FT DOMAIN 131 143
CC FT POLY-HIS.
CC FT DOMAIN 158 167
CC FT POLY-HIS.
CC FT DOMAIN 194 201
CC FT POLY-ALA.
CC SQ SEQUENCE 323 AA; 35792 MW; 6E386340D1F840A5 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 323;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10
DB 57 STPLSTPC 64

RESULT 14
V181_FOWPV
ID V181_FOWPV STANDARD; PRT; 369 AA.
AC Q9J552;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein FV181.
GN FV181.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;
RA Afonso C.L., Tullman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831(2000).
CC -!- SIMILARITY: Belongs to the poxviruses A16 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF198100; AAF44525.1; -.
CC InterPro; IPR004251; DUF230.

```

```
DR Pfam: PF03003; DUF230; 1.
SQ SEQUENCE 369 AA; 42081 MW; 380A71032C18BB99 CRC64;

Query Match      62.5%; Score 35; DB 1; Length 369;
Best Local Similarity 50.0%; Pred.No. 50;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 10
   |  |  |  |  |  |
Db 164 CLQTGSLPC 173

RESULT 15
FD21_MYCTU STANDARD; PRT; 578 AA.
ID AC OS0441;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative fatty-acid-CoA ligase fadD21 (EC 6.2.1.-) (Acyl-CoA
DE synthetase).
GN FADD21 OR RV1185C OR MT1222 OR MTV005.21C OR MB1217C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=12218036;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Winn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RA J. Bacteriol. 184:5479-5490(2002).
RN [3]
RN SEQUENCE FROM N.A.
RP SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC
```

```
CC or send an email to licenses@isb-sib.ch.
CC -----
CC EMBL; AL010186; CAA15862.1; -
CC EMBL; AB006999; AAK45479.1; -
CC EMBL; BX248338; CAD94078.1; -
CC FIR; A70877; A70877.
CC TIGR; MT1222; -
CC TubercuList; RV1185C; -
CC InterPro: IPR000873; AMP-bind.
CC Pfam; PF00501; AMP-binding; 1.
CC PROSITE; PS00455; AMP_BINDING; FALSE_NEG.
CC KW Hypothetical protein; Ligase; Fatty acid metabolism;
CC Complete proteome.
CC SQ SEQUENCE 578 AA; 62756 MW; 698FE8FE9E4BFA2F CRC64;

Query Match      62.5%; Score 35; DB 1; Length 578;
Best Local Similarity 87.5%; Pred.No. 79;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ISVPLSVP 9
   |  |  |  |  |  |
Db 88 IAVPLSVP 95

Search completed: September 5, 2004, 09:56:10
Job time : 5.54545 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:46:09 ; Search time 22.6263 Seconds  
(without alignments)  
139.448 Million cell updates/sec

Title: US-09-761-636A-13  
Perfect score: 56  
Sequence: 1 CISVPLSVPC 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	40	71.4	109	2	Q914Y0	streptomyces
2	39	69.6	85	2	Q9S556	pseudomonas
3	39	69.6	85	16	Q9S564	pseudomonas
4	39	69.6	148	16	Q9PBM2	Q9pbm2 xylella fas
5	39	69.6	620	12	Q98983	Q98983 human herpe
6	38	67.9	144	3	Q8TGP8	Q8tgp8 saccharomyc
7	38	67.9	503	10	Q8L8U2	Q8l8u2 arabidopsis
8	38	67.9	528	16	Q89DL6	Q89dl6 bradyrhizob
9	38	67.9	563	16	Q88SR1	Q88sr1 lactobacill
10	37	66.1	86	9	Q9MC88	Q9mc88 bacterioph
11	37	66.1	89	16	Q9P9Y4	Q9p9y4 xylella fas
12	37	66.1	129	11	Q8BWA6	Q8bwa6 mus musculu
13	37	66.1	134	4	Q8N7T2	Q8n7t2 homo sapien
14	37	66.1	155	16	Q8DMP5	Q8dmp5 synectococc
15	37	66.1	236	16	Q87XN8	Q87xn8 pseudomonas
16	37	66.1	284	16	Q92RN5	Q92rn5 rhizobium m

17	37	66.1	316	5	Q9V743	Q9v743 drosophila
18	37	66.1	355	10	Q9LGY9	Q9lgy9 oryza sativ
19	37	66.1	370	5	Q9VA44	Q9va44 drosophila
20	37	66.1	501	10	Q9AY75	Q9ay75 oryza sativ
21	37	66.1	725	5	Q9N8V1	Q9n8v1 trypanosoma
22	36	64.3	58	16	Q7URU5	Q7urus rhodopirell
23	36	64.3	112	4	Q14634	Q14634 homo sapien
24	36	64.3	134	11	Q63318	Q63318 rattus norv
25	36	64.3	207	16	Q9CG60	Q9c960 lactococcus
26	36	64.3	208	5	Q9NF93	Q9nf93 leishmania
27	36	64.3	256	16	Q9RX44	Q9rx44 deinococcus
28	36	64.3	280	10	Q94JK8	Q94jk8 oryza sativ
29	36	64.3	496	16	Q9HMV6	Q9hmw6 pseudomonas
30	36	64.3	512	10	Q9FX29	Q9fx29 arabidopsis
31	36	64.3	668	4	Q9BR66	Q9br66 homo sapien
32	36	64.3	857	10	Q9ASK9	Q9ask9 oryza sativ
33	36	64.3	1016	11	Q8C310	Q8c310 mus musculu
34	36	64.3	1136	13	Q98UH3	Q98uh3 xenopus lae
35	35.5	63.4	1904	13	Q8AV28	Q8av28 gallus gall
36	35	62.5	77	3	Q87ZN3	Q87zn3 neurospora
37	35	62.5	100	17	Q9YG98	Q9yg98 aeropyrum p
38	35	62.5	114	11	Q8CCF3	Q8ccf3 mus musculu
39	35	62.5	133	3	Q870B4	Q870b4 piromyces s
40	35	62.5	133	4	Q8N218	Q8n218 homo sapien
41	35	62.5	144	4	Q9BZ14	Q9bz14 homo sapien
42	35	62.5	161	5	Q9NB94	Q9nb94 anopheles s
43	35	62.5	192	4	Q9BZ15	Q9bz15 homo sapien
44	35	62.5	203	4	Q8WY4	Q8wy4 homo sapien
45	35	62.5	211	10	Q8L4L5	Q8l4l5 oryza sativ

## ALIGNMENTS

## RESULT 1

Q914Y0 PRELIMINARY; PRT; 109 AA.  
ID Q914Y0  
AC Q914Y0  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Tcm F1 monoxygenase.  
GN ELMH  
OS Streptomyces olivaceus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=47716;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tu 2353;  
RX MEDLINE=21225632; PubMed=11325225;  
RA Rafanan E.R. Jr., Le L., Zhao L., Decker H., Shen B.;  
RT "Cloning, sequencing, and heterologous expression of the elmGH1J genes  
involved in the biosynthesis of the polyketide antibiotic elloramycin  
from Streptomyces olivaceus Tu2353.";  
RL J. Nat. Prod. 64:444-449(2001).  
DR EMBL; AF263463; AAF73051.1; -.  
DR GO; GO:0004497; F:monoxygenase activity; IEA.  
DR InterPro; IPR007138; ABM.  
DR Pfam; PF03992; ABM; 1.  
KW Monoxygenase.  
SQ SEQUENCE 109 AA; 12304 MW; E65F4010C7D4B30B CRC64;

Query Match 71.4%; Score 40; DB 2; Length 109;  
Best Local Similarity 77.8%; Pred. No. 7.3;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISVPLSVPC 10  
|:|||||  
Db 90 IATPLSVPC 98

## RESULT 2

"Complete genome sequence of Pseudomonas aeruginosa PA01, an RT opportunistic pathogen.";									
RT	Q9S556	PRELIMINARY;	PRT;	85 AA.					
AC	Q9S556;								
DR	EMBL; AB010825; BAA83170.1; -.								
DR	EMBL; AB045308; BAA97449.1; -.								
DR	EMBL; AE004498; BAA04020.1; -.								
DR	PIR; T44555; T44555.								
DR	InterPro; IPR000437; Prok_lipoprot S.								
DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.								
KW	Hypothetical protein; Complete proteome.								
SQ	SEQUENCE 85 AA; 9222 MW; F2267D9C542C7F48 CRC64;								
Query Match 69.6%; Score 39; DB 16; Length 85;									
Best Local Similarity 66.7%; Pred. No. 8.8;									
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;									
QY	2	ISVPLSVPC	10						
	:	:							
DB	27	VEVPLAVPC	35						
RESULT 4									
ID	Q9PBM2	PRELIMINARY;	PRT;	148 AA.					
AC	Q9PBM2;								
DT	01-OCT-2000 (TrEMBLrel. 15, Created)								
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)								
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)								
DE	Hypothetical protein Xf2118.								
GN	Xf2118.								
OS	Xylella fastidiosa.								
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;								
OC	Xanthomonadaceae; Xylella.								
OX	NCBI_TaxID=2371;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=9a5C;								
RX	MEDLINE=20365717; PubMed=10910347;								
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,								
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,								
RA	Bartos M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,								
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,								
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,								
RA	Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,								
RA	Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferri J.A.,								
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,								
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,								
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,								
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,								
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,								
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,								
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,								
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,								
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,								
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,								
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,								
RA	Peixoto R.B., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,								
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,								
RA	da Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,								
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,								
RA	da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,								
RA	de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,								
RA	Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,								
RA	Zago M.A., Zatz M., Meidanis J., Setubal J.C.;								
RT	"The genome sequence of the plant pathogen Xylella fastidiosa.";								
RL	Nature 406:151-159 (2000).								
DR	EMBL; AE004026; AAF84917.1; -.								
DR	PIR; G82599; G82599.								
KW	Hypothetical protein; Complete proteome.								
SQ	SEQUENCE 148 AA; 15990 MW; B78D2FAE6B929260 CRC64;								
Query Match 69.6%; Score 39; DB 16; Length 148;									
Best Local Similarity 50.0%; Pred. No. 15;									

"Complete genome sequence of Pseudomonas aeruginosa PA01, an RT opportunistic pathogen.";									
RT	Q9S556	PRELIMINARY;	PRT;	85 AA.					
AC	Q9S556;								
DR	EMBL; AB010825; BAA83170.1; -.								
DR	EMBL; AB045308; BAA97449.1; -.								
DR	EMBL; AE004498; BAA04020.1; -.								
DR	PIR; T44555; T44555.								
DR	InterPro; IPR000437; Prok_lipoprot S.								
DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.								
KW	Hypothetical protein; Complete proteome.								
SQ	SEQUENCE 85 AA; 9222 MW; F2267D9C542C7F48 CRC64;								
Query Match 69.6%; Score 39; DB 16; Length 85;									
Best Local Similarity 66.7%; Pred. No. 8.8;									
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;									
QY	2	ISVPLSVPC	10						
	:	:							
DB	27	VEVPLAVPC	35						
RESULT 3									
ID	Q9S564	PRELIMINARY;	PRT;	85 AA.					
AC	Q9S564;								
DT	01-MAY-2000 (TrEMBLrel. 13, Created)								
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)								
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)								
DE	Hypothetical protein PA0631.								
GN	PA0631.								
OS	Pseudomonas aeruginosa.								
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;								
OC	Pseudomonadaceae; Pseudomonas.								
OX	NCBI_TaxID=287;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=ATCC 15692 / PA01;								
RA	Nakayama K., Takashima K., Ishihara H., Shinomiya T., Kageyama M.,								
RA	Kanaya S., Ohnishi M., Murata T., Terawaki Y., Mori H., Hayashi T.;								
RT	"Genetic relationship between bacteriocins and bacteriophages.";								
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=SLM6;								
RA	Nakayama K., Takashima K., Ishihara H., Shinomiya S., Kageyama M.,								
RA	Kanaya S., Ohnishi M., Murata T., Mori H., Hayashi T.;								
RT	"The R-type pyocin is related to P2 phage, and the F-type pyocin is								
RT	related to lambda phage.";								
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.								
RN	[3]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=ATCC 15692 / PA01;								
RX	MEDLINE=20437337; PubMed=10984043;								
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,								
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,								
RA	Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,								
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,								
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,								
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;								
Query Match 69.6%; Score 39; DB 16; Length 148;									
Best Local Similarity 50.0%; Pred. No. 15;									

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 C1SVPLSVC 10  
| :||: ||  
Db 8 CCAVPMNAPC 17

RESULT 5  
Q89893 PRELIMINARY; PRT; 620 AA.

AC Q89893; 1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN DR2.  
OS Human herpesvirus 6.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Roseolovirus.  
OX NCBI\_TaxID=10368;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=95266321; PubMed=7747482;  
RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,  
RA Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;  
RT "The DNA sequence of human herpesvirus-6: structure, coding content,  
RT and genome evolution.";  
RL Virology 209:29-51(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=90080132; PubMed=2152817;  
RA Lawrence G.L., Chee M., Craxton M.A., Gompels U.A., Honess R.W.,  
RA Barrell B.G.;  
RT "Human herpesvirus 6 is closely related to human cytomegalovirus.";  
RL J. Virol. 64:287-299(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=91237802; PubMed=1851860;  
RA Chang C.K., Balachandran N.;  
RT "Identification, characterization, and sequence analysis of a cDNA  
RT encoding a phosphoprotein of human herpesvirus 6.";  
RL J. Virol. 65:2884-2894(1991).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=91333007; PubMed=1651403;  
RA Teo I.A., Griffin B.E., Jones M.D.;  
RT "Characterization of the DNA polymerase gene of human herpesvirus 6.";  
RL J. Virol. 65:4670-4680(1991).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=91226542; PubMed=1851252;  
RA Thomson B.J., Efstathiou S., Honess R.W.;  
RT "Acquisition of the human adeno-associated virus type-2 rep gene by  
RT human herpesvirus type-6.";  
RL Nature 351:78-80(1991).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=91374590; PubMed=1654446;  
RA Martin M.E., Nicholas J., Newman C., Honess R.W.;  
RT "Identification of a transactivating function mapping to the putative  
RT immediate-early locus of human herpesvirus 6.";  
RL J. Virol. 65:5381-5390(1991).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=92333249; PubMed=1321206;  
RA Efstathiou S., Lawrence G.L., Brown C.M., Barrell B.G.;

RT "Identification of homologues to the human cytomegalovirus US22 gene  
RT family in human herpesvirus 6.";  
RL J. Gen. Virol. 73:1661-1671(1992).  
RN [8]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=92148942; PubMed=1310766;  
RA Gang Y., Chandran B., Josephs S.F., Wood C.;  
RT "Identification and characterization of a human herpesvirus 6 gene  
RT segment that trans activates the human immunodeficiency virus type 1  
RT promoter.";  
RL J. Virol. 66:1564-1570(1992).  
RN [9]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=93091236; PubMed=1333836;  
RA Gompels U.A., Carss A.L., Sun N., Arrand J.R.;  
RT "Infectivity determinants encoded in a conserved gene block of human  
RT herpesvirus-6.";  
RL DNA Seq. 3:25-39(1992).  
RN [10]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=92260671; PubMed=1374813;  
RA Neipel F., Ellinger K., Fleckenstein B.;  
RT "Gene for the major antigenic structural protein (p100) of human  
RT herpesvirus 6.";  
RL J. Virol. 66:3918-3924(1992).  
RN [11]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=92333248; PubMed=1321205;  
RA Thomson B.J., Honess R.W.;  
RT "The right end of the unique region of the genome of human herpesvirus  
RT 6 U1102 contains a candidate immediate early gene enhancer and a  
RT homologue of the human cytomegalovirus US22 gene family.";  
RL J. Gen. Virol. 73:1649-1660(1992).  
RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=93187613; PubMed=8383182;  
RA Ellinger K., Neipel F., Foa-Tomasi L., Campadelli-Piome G.,  
RA Fleckenstein B.;  
RT "The glycoprotein B homologue of human herpesvirus 6.";  
RL J. Gen. Virol. 74:495-500(1993).  
RN [13]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=93224882; PubMed=8385692;  
RA Gompels U.A., Carrigan D.R., Carss A.L., Arno J.;  
RT "Two groups of human herpesvirus 6 identified by sequence analyses of  
RT laboratory strains and variants from Hodgkin's lymphoma and bone  
RT marrow transplant patients.";  
RL J. Gen. Virol. 74:613-622(1993).  
RN [14]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=93389439; PubMed=8397282;  
RA Liu D.X., Gompels U.A., Nicholas J., Lelliott C.;  
RT "Identification and expression of the human herpesvirus 6 glycoprotein  
RT H and interaction with an accessory 40K glycoprotein.";  
RL J. Gen. Virol. 74:1847-1857(1993).  
RN [15]  
RP SEQUENCE FROM N.A.  
RC STRAIN=U1102;  
RX MEDLINE=94025558; PubMed=7692666;  
RA Liu D.X., Gompels U.A., Foa-Tomasi L., Campadelli-Piome G.;  
RT "Human herpesvirus-6 glycoprotein H and L homologs are components of  
RT the gp100 complex and the gp100 external domain is the target for  
RT neutralizing monoclonal antibodies.";  
RL Virology 197:12-22(1993).  
RN [16]  
RP SEQUENCE FROM N.A.





```

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPLSVPC 10
Db 49 VPLSIPC 55

RESULT 8
Q89D16
ID Q89D16 PRELIMINARY; PRT; 528 AA.
AC Q89D16;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE BLr7629 protein.
GN BLR7629.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005962; BAC52894.1; -
DR InterPro; IPR002830; carboxylase.
DR PFam; PF01977; Ubid. 1
DR PROSITE; PS00024; HEMOPEXIN; 1.
KW Complete proteome.
SQ SEQUENCE 528 AA; 58441 MW; 4A1DB04BD1B08546 CRC64;

Query Match 67.9%; Score 38; DB 16; Length 528;
Best Local Similarity 66.7%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CISVPLSVP 9
Db 284 CLTVPLEVP 292

RESULT 9
Q88SR1
ID Q88SR1 PRELIMINARY; PRT; 563 AA.
AC Q88SR1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Adenine deaminase (EC 3.5.4.2).
GN ADEC OR LP 3334.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Turchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935261; CAD65449.1; -

```

```

DR GO; GO:0000034; F:adenine deaminase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR006680; Amidohydro_1.
DR Pfam; PF01979; Amidohydro_1; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 563 AA; 60927 MW; A4B5159C11718FA7 CRC64;

Query Match 67.9%; Score 38; DB 16; Length 563;
Best Local Similarity 60.0%; Pred. No. 81;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CISVPLSVP 10
Db 120 CFMLPSSVPC 129

RESULT 10
Q9MC88
ID Q9MC88 PRELIMINARY; PRT; 86 AA.
AC Q9MC88;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Orf33.
GN ORF33.
OS Bacteriophage D3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=31535;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20042341; PubMed=10572124;
RA Gilakjan Z.A., Kropinski A.M.;
RT "Cloning and analysis of the capsid morphogenesis genes of Pseudomonas
RT aeruginosa bacteriophage D3: another example of protein chain mail?";
RL J. Bacteriol. 181:7221-7227(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20485557; PubMed=11029426;
RA Kropinski A.M.;
RT "Sequence of the Genome of the Temperate, Serotype-Converting,
RT Pseudomonas aeruginosa Bacteriophage D3.";
RL J. Bacteriol. 182:6066-6074(2000).
DR EMBL; AF165214; AAF80793.1; -
DR InterPro; IPR000437; Prok_lipoprot_S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
SQ SEQUENCE 86 AA; 9389 MW; 86ABCAB2B2A9E59F CRC64;

Query Match 66.1%; Score 37; DB 9; Length 86;
Best Local Similarity 55.6%; Pred. No. 21;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISVPLSVP 10
Db 28 VEVPIAVPC 36

RESULT 11
Q9P9Y4
ID Q9P9Y4 PRELIMINARY; PRT; 89 AA.
AC Q9P9Y4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Hypothetical protein Xf2736.
GN Xf2736.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;

```

```
RX MEDLINE-20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barro M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lenos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR PIR; F82521;
DR EMBL; AE004080; AAF85521.1; -.
SQ SEQUENCE 89 AA; 9641 MW; F13175F1054D6130 CRC64;
Query Match 66.1%; Score 37; DB 16; Length 89;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 ISVPLSVPC 10
Db 19 VSAPLRVPC 27
PRELIMINARY; PRT; 129 AA.
AC Q8BWA6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical sugar phosphatases structure containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053082; BAC35260.1; -.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13584 MW; D0B83B75AFAD595B CRC64;
Query Match 66.1%; Score 37; DB 11; Length 129;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 ISVPLSVPC 9
Db 19 VSAPLRVPC 27
PRELIMINARY; PRT; 129 AA.
AC Q8BWA6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical sugar phosphatases structure containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053082; BAC35260.1; -.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13584 MW; D0B83B75AFAD595B CRC64;
Query Match 66.1%; Score 37; DB 16; Length 89;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 ISVPLSVPC 10
Db 19 VSAPLRVPC 27
PRELIMINARY; PRT; 129 AA.
AC Q8BWA6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical sugar phosphatases structure containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053082; BAC35260.1; -.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13584 MW; D0B83B75AFAD595B CRC64;
Query Match 66.1%; Score 37; DB 16; Length 155;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CISVPLSVPC 9
Db 46 CIAGPLLLPC 55
PRELIMINARY; PRT; 155 AA.
AC Q8DMP5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TLR0066 protein.
GN TLR0066.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_taxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005369; BAC07619.1; -.
KW Complete proteome.
SQ SEQUENCE 155 AA; 17833 MW; 16CE9F9F696CEC331 CRC64;
Query Match 66.1%; Score 37; DB 16; Length 155;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CISVPLSVPC 9
Db 46 CIAGPLLLPC 55
PRELIMINARY; PRT; 155 AA.
AC Q8DMP5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TLR0066 protein.
GN TLR0066.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_taxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005369; BAC07619.1; -.
KW Complete proteome.
SQ SEQUENCE 155 AA; 17833 MW; 16CE9F9F696CEC331 CRC64;
Query Match 66.1%; Score 37; DB 4; Length 134;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CISVPLSVPC 10
Db 46 CIAGPLLLPC 55
PRELIMINARY; PRT; 134 AA.
AC Q8N7T2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ40382.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097701; BAC05145.1; -.
KW Hypothetical protein.
SQ SEQUENCE 134 AA; 14294 MW; 06EA299087C8C41F CRC64;
Query Match 66.1%; Score 37; DB 4; Length 134;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CISVPLSVPC 10
Db 46 CIAGPLLLPC 55
PRELIMINARY; PRT; 134 AA.
AC Q8N7T2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ40382.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097701; BAC05145.1; -.
KW Hypothetical protein.
SQ SEQUENCE 134 AA; 14294 MW; 06EA299087C8C41F CRC64;
```

```
RX MEDLINE-20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barro M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lenos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR PIR; F82521;
DR EMBL; AE004080; AAF85521.1; -.
SQ SEQUENCE 89 AA; 9641 MW; F13175F1054D6130 CRC64;
Query Match 66.1%; Score 37; DB 16; Length 89;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 ISVPLSVPC 10
Db 19 VSAPLRVPC 27
PRELIMINARY; PRT; 129 AA.
AC Q8BWA6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical sugar phosphatases structure containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053082; BAC35260.1; -.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13584 MW; D0B83B75AFAD595B CRC64;
Query Match 66.1%; Score 37; DB 11; Length 129;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 ISVPLSVPC 9
Db 19 VSAPLRVPC 27
PRELIMINARY; PRT; 129 AA.
AC Q8BWA6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical sugar phosphatases structure containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053082; BAC35260.1; -.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13584 MW; D0B83B75AFAD595B CRC64;
Query Match 66.1%; Score 37; DB 16; Length 89;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 2 ISVPLSVPC 10
Db 19 VSAPLRVPC 27
PRELIMINARY; PRT; 129 AA.
AC Q8BWA6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical sugar phosphatases structure containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK053082; BAC35260.1; -.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13584 MW; D0B83B75AFAD595B CRC64;
Query Match 66.1%; Score 37; DB 16; Length 155;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CISVPLSVPC 9
Db 46 CIAGPLLLPC 55
PRELIMINARY; PRT; 155 AA.
AC Q8DMP5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TLR0066 protein.
GN TLR0066.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_taxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005369; BAC07619.1; -.
KW Complete proteome.
SQ SEQUENCE 155 AA; 17833 MW; 16CE9F9F696CEC331 CRC64;
Query Match 66.1%; Score 37; DB 4; Length 134;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CISVPLSVPC 10
Db 46 CIAGPLLLPC 55
PRELIMINARY; PRT; 134 AA.
AC Q8N7T2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ40382.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097701; BAC05145.1; -.
KW Hypothetical protein.
SQ SEQUENCE 134 AA; 14294 MW; 06EA299087C8C41F CRC64;
```

Db 94 CQSLPLSV 102

## RESULT 15

Q87XN8 PRELIMINARY; PRT; 236 AA.  
 ID Q87XN8  
 AC Q87XN8;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Amino acid ABC transporter, permease protein.  
 GN PSPT04138.  
 OS Pseudomonas syringae (pv. tomato).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
 RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,  
 RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
 RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,  
 RA White O., Fraser C., Collmer A.;  
 RT "Complete sequence of Pseudomonas syringae."  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016870; AAC057594.1; -  
 DR TIGR; PSPT04138; -  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR000515; BPD transp.  
 DR Pfam; PF00528; BPD transp. 1.  
 DR PROSITE; PS00402; BPD\_TRANS\_INN\_MEMBER; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 236 AA; 26706 MW; 05A989893BFF33DC CRC64;

Query Match 66.1%; Score 37; DB 16; Length 236;  
 Best Local Similarity 62.5%; Pred. No. 54;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQSVPLSV 8

Db 39 CQSVPLSI 46

Search completed: September 5, 2004, 10:00:07  
 Job time : 24.6263 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:37:49 ; Search time 32.8283 Seconds  
(without alignments)  
86.068 Million cell updates/sec

Title: US-09-761-636A-13  
Perfect score: 56  
Sequence: 1 CTSVPLSVPC 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	56	100.0	10	4	AAU04532	AAU04532 VEGF base
2	45.5	81.2	11	4	AAU04526	AAU04526 VEGF base
3	45	80.4	141	2	AAR66394	AAR66394 Human psy
4	43.5	77.7	11	4	AAU04542	AAU04542 VEGF base
5	41.5	74.1	9	4	AAU04533	AAU04533 VEGF base
6	41.5	74.1	11	4	AAU04545	AAU04545 VEGF base
7	40.5	72.3	11	4	AAU04544	AAU04544 VEGF base
8	39	69.6	77	4	AAU95534	AAU95534 Human rep
9	39	69.6	77	4	ABB96216	ABB96216 Human tes
10	39	69.6	108	4	AAU95827	AAU95827 Human rep
11	39	69.6	108	4	ABB96358	ABB96358 Human tes
12	38.5	68.8	11	4	AAU04543	AAU04543 VEGF base
13	38	67.9	59	4	ABG12542	ABG12542 Novel hum
14	38	67.9	59	4	ABG12530	ABG12530 Novel hum
15	38	67.9	71	4	AAU16003	AAU16003 Peptide #
16	38	67.9	71	4	ABB20407	ABB20407 Protein #
17	38	67.9	71	5	ABG37710	ABG37710 Human rep
18	38	67.9	217	7	ADB64084	ADB64084 Human pro
19	38	67.9	502	3	AAG30888	AAG30888 Arabidops
20	38	67.9	503	3	AAG20989	AAG20989 Arabidops
21	38	67.9	503	3	AAG43108	AAG43108 Arabidops
22	37	66.1	108	4	ABB15040	ABB15040 Human ner
23	37	66.1	316	4	ABB67498	ABB67498 Drosophil
24	37	66.1	368	4	ABG21894	ABG21894 Novel hum
25	37	66.1	370	4	ABB65635	ABB65635 Drosophil

## ALIGNMENTS

RESULT 1  
AAU04532  
ID AAU04532 standard; peptide; 10 AA.  
XX AC AAU04532;  
XX AC  
DT 26-SEP-2001 (first entry)  
XX  
DE VEGF based monocyclic peptide 10.  
XX  
KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
KW diabetic retinopathy; chronic inflammation; cyclic.  
XX Synthetic.  
OS  
XX Key Location/Qualifiers  
FH Disulfide-bond 1..10  
FT /note= "This bond cyclises the peptide"  
FT  
XX  
PN WO200152875-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 18-JAN-2001; 2001WO-US001533.  
XX  
PR 18-JAN-2000; 2000US-0176293P.  
PR 16-MAY-2000; 2000US-0204590P.  
XX (LUDW-) LUDWIG INST CANCER RES.  
PA Achen MG, Hughes RA, Stackler S, Cendron A;  
XX WPI; 2001-442248/47.  
XX  
PT Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
PT or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
PT from an exposed loop of a growth factor protein by oxidizing the cysteine  
PT residues.  
XX Claim 49; Page 32; 102pp; English.  
XX  
CC The sequence represents a monomeric monocyclic peptide of the invention,  
CC whose 3-dimensional structure is modelled on the expose loop of human  
CC VEGF (vascular endothelial growth factor). The invention relates to a  
CC method of producing a monomeric monocyclic peptide by a measuring beta-  
CC beta carbon separation distances on opposite antiparallel strands of a

Abg20151 Novel hum  
Abg71244 Guinea pi  
Aar96138 Protease  
Aaw82212 D-NorFES-  
Aaw45662 Peptide b  
Aau58478 Propionib  
Abm54997 Propionib  
Abp64007 Human ORP  
Aab45459 Human sec  
Aab45460 Human sec  
Aag89148 Human sec  
Aab45461 Human sec  
Aab45462 Human sec  
Abg19452 Novel hum  
Aau66053 Propionib  
Abm62572 Propionib  
Aag04076 Human sec  
Aam79027 Human pro  
Aam80011 Human pro  
Abb54586 Lactococc

CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis,  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 SQ Sequence 10 AA;

Query Match 100.0%; Score 56; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.0098;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CISVPLSVPC 10  
 |||||  
 Db 1 CISVPLSVPC 10

RESULT 2  
 AAU04526 1.11  
 ID AAU04526 standard; peptide; 11 AA.

XX AC AAU04526;  
 XX DT 26-SEP-2001 (first entry)  
 XX DE VEGF based monocyclic peptide 3.  
 XX KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
 KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
 KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
 KW diabetic retinopathy; chronic inflammation; cyclic.  
 XX OS Synthetic.

XX FH Key Location/Qualifiers  
 FT Disulfide-bond 1.11  
 FT /note= "This bond cyclises the peptide"

XX PN WO200152875-A1.  
 XX PD 26-JUL-2001.  
 XX XX 18-JAN-2001; 2001WO-US001533.  
 XX XX 18-JAN-2000; 2000US-0176293P.  
 XX PR 16-MAY-2000; 2000US-0204590P.  
 XX XX (LUDW-) LUDWIG INST CANCER RES.  
 XX PA Achen MG, Hughes RA, Stacker S, Cendron A;  
 XX FI WPI; 2001-442248/47.  
 XX DR Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
 XX PT or lymphangiogenesis, is produced by cyclising a peptide loop fragment  
 XX PT from an exposed loop of a growth factor protein by oxidizing the cysteine

PT residues.  
 XX Claim 49; Page 32; 102pp; English.  
 XX The sequence represents a monomeric monocyclic peptide of the invention,  
 CC whose 3-dimensional structure is modelled on the expose loop of human  
 CC VEGFD (vascular endothelial growth factor). The invention relates to a  
 CC method of producing a monomeric monocyclic peptide by a measuring beta-  
 CC beta carbon separation distances on opposite antiparallel strands of a  
 CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis,  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 SQ Sequence 11 AA;

Query Match 81.2%; Score 45.5; DB 4; Length 11;  
 Best Local Similarity 90.9%; Pred. No. 0.56;  
 Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Qy 1 CISVPL-SVPC 10  
 |||||  
 Db 1 CISVPLTSVPC 11

RESULT 3  
 AAR66394  
 ID AAR66394 standard; peptide; 141 AA.  
 XX AC AAR66394;  
 XX DT 25-MAR-2003 (revised)  
 DT 11-AUG-1995 (first entry)  
 XX DE Human psychosis protecting peptide encoded by reading frame 1.  
 XX KW Psychosis protecting peptide.  
 XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 FT Misc-difference 1.141  
 FT /note= "X corresp. to translated stop codon"  
 XX PN WO9426107-A1.  
 XX PD 24-NOV-1994.  
 XX XX 13-MAY-1994; 94WO-US005445.  
 XX PR 13-MAY-1993; 93US-00060560.  
 XX XX (UYNY ) UNIV NEW YORK STATE.  
 XX PA Friedhoff AJ, Basham DA, Miller JC;  
 XX PI

XX WPI; 1995-006234/01.  
 DR N-PSDB; AAQ75090.  
 XX  
 XX New nucleic acids encoding psychosis protecting peptide and antibodies -  
 PT for the treatment, diagnosis and research of psychotic disorders, such as  
 PT schizophrenia.  
 XX  
 XX Claim 1; Page 56; 87pp; English.  
 PS  
 XX  
 CC Psychosis protecting peptide (PP peptide) can be obtd. initially by  
 CC using the sequence in AAQ75090 as a basis for designing polynucleotide  
 CC probes to clone, sequence and express or synthesize PP related protein  
 CC and peptides occurring in normal individuals, and to a substantially  
 CC lesser degree in individuals with psychotic disorders. The PP encoding  
 CC gene was discovered using the subtraction cloning of cDNA from mRNA obtd.  
 CC from monozygotic twins discordant for schizophrenia and assaying for  
 CC clones in which expression is greatest in the 'well' twin. A nt sequence  
 CC substantially corresp. to AAQ75090 is claimed, which comprises 30-X nts,  
 CC where X=80, 95, 158, 222, 243, 249, 260, 295, 407 or 423. A PP peptide of  
 CC 10-141 AAs is also claimed. (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 141 AA;  
 Query Match 80.4%; Score 45; DB 2; Length 141;  
 Best Local Similarity 77.8%; Pred. No. 9.4;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTSVPLSVP 9  
 DB 28 CTSIPLSIP 36  
 RESULT 4  
 AAU04542  
 ID AAU04542 standard; peptide; 11 AA.  
 XX  
 AC AAU04542;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE VEGF based monocyclic peptide 20.  
 XX  
 KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
 KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
 KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
 KW diabetic retinopathy; chronic inflammation; cyclic.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 1..11 /note= "This bond cyclises the peptide"  
 FT  
 XX  
 XX WO200152875-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 18-JAN-2001; 2001WO-US001533..  
 XX  
 XX 18-JAN-2000; 2000US-0176293P.  
 PR 16-MAY-2000; 2000US-0204590P.  
 XX  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 PA  
 XX Achen MG, Hughes RA, Stacker S, Cendron A;  
 PI WPI; 2001-442248/47.  
 XX  
 XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
 PT or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
 PT from an exposed loop of a growth factor protein by oxidizing the cysteine  
 PT residues.

XX  
 PS  
 XX  
 CC Example 25; Page 47; 102pp; English.  
 CC  
 CC The sequence represents a monomeric monocyclic peptide of the invention,  
 CC whose 3-dimensional structure is modelled on the expose loop of human  
 CC VEGFD (vascular endothelial growth factor). The invention relates to a  
 CC method of producing a monomeric monocyclic peptide by a measuring beta-  
 CC beta carbon separation distances on opposite antiparallel strands of a  
 CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis.  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 SQ Sequence 11 AA;  
 Query Match 77.7%; Score 43.5; DB 4; Length 11;  
 Best Local Similarity 81.8%; Pred. No. 1.2;  
 Matches 9; Conservative 1; Mismatches 0; Indels 1; Gaps 1;  
 QY 1 CTSVPL-SVPC 10  
 DB 1 CLSVPLTSVPC 11  
 RESULT 5  
 AAU04533  
 ID AAU04533 standard; peptide; 9 AA.  
 XX  
 AC AAU04533;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE VEGF based monocyclic peptide 11.  
 XX  
 KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
 KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
 KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
 KW diabetic retinopathy; chronic inflammation; cyclic.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 1..9 /note= "This bond cyclises the peptide"  
 FT  
 XX  
 XX WO200152875-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 18-JAN-2001; 2001WO-US001533.  
 XX  
 XX 18-JAN-2000; 2000US-0176293P.  
 PR 16-MAY-2000; 2000US-0204590P.  
 XX  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 PA

```

XX PI Achen MG, Hughes RA, Stacker S, Cendron A;
XX DR WPI; 2001-442248/47.
XX
XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,
XX or lymphangiogenesis, is produced by cyclizing a peptide loop fragment
XX from an exposed loop of a growth factor protein by oxidizing the cysteine
XX residues.
XX
XX Claim 49; Page 32; 102pp; English.
XX
XX The sequence represents a monomeric monocyclic peptide of the invention,
XX whose 3-dimensional structure is modelled on the expose loop of human
XX VEGF (vascular endothelial growth factor). The invention relates to a
XX method of producing a monomeric monocyclic peptide by a measuring beta-
XX beta carbon separation distances on opposite antiparallel strands of a
XX peptide loop fragment from an exposed loop of a growth factor protein and
XX cyclising the peptide by oxidising the cysteine residues. The monocyclic
XX peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic
XX peptides) and a cyclic peptide with at least one amino acid deleted prior
XX to cyclisation are used to interfere with angiogenesis,
XX neovascularisation or lymphangiogenesis in a mammal with a condition
XX characterised by angiogenesis, neovascularisation or lymphangiogenesis.
XX The condition is diabetic retinopathy, psoriasis, arthropathy,
XX hemangioma, vascularised malignant or benign tumour, post-recovery
XX cerebrovascular accident, post-angioplasty restenosis, head, heat or cold
XX trauma, substance-induced neovascularisation of the liver, excessive
XX hormone-related angiogenic dysfunction, diabetes induced neovascular
XX sequelae, hypertension induced neovascular sequelae, or chronic liver
XX infection. The peptides are also used to modulate vascular permeability
XX in a mammal (the mammal has a condition characterised by fluid
XX accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,
XX or brain. The peptides are used to image blood vessels and lymphatic
XX vasculature. The monomeric and bicyclic peptides are used to interfere
XX with at least one biological activity induced by VEGF, VEGF-C or -D and
XX are also used in combination with an anti-inflammatory agent, to treat a
XX chronic inflammation, especially rheumatoid arthritis, psoriasis and
XX diabetic retinopathy
XX
XX Sequence 9 AA;
XX
XX Query Match 74.1%; Score 41.5; DB 4; Length 9;
XX Best Local Similarity 90.0%; Pred. No. 1.4e+06;
XX Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
XX
XX Qy 1 CISVPLSVPC 10
XX ||||| |||
XX 1 CISVPL-VPC 9
XX
XX RESULT 6
XX AAU04545
XX ID AAU04545 standard; peptide; 11 AA.
XX
XX AC AAU04545;
XX
XX DT 26-SEP-2001 (first entry)
XX
XX DE VEGF based monocyclic peptide 23.
XX
XX Human; VEGF; vascular endothelial growth factor; angiogenesis;
XX neovascularisation; lymphangiogenesis; psoriasis; tumour;
XX diabetes induced neovascular sequelae; rheumatoid arthritis;
XX diabetic retinopathy; chronic inflammation; cyclic.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 1..11
XX /note= "This bond cyclises the peptide"
XX
XX WO200152875-A1.

```

```

XX PD 26-JUL-2001.
XX PF 18-JAN-2001; 2001WO-US001533.
XX
XX PR 18-JAN-2000; 2000US-0176293P.
XX PR 16-MAY-2000; 2000US-0204590P.
XX
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX
XX PI Achen MG, Hughes RA, Stacker S, Cendron A;
XX DR WPI; 2001-442248/47.
XX
XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,
XX or lymphangiogenesis, is produced by cyclizing a peptide loop fragment
XX from an exposed loop of a growth factor protein by oxidizing the cysteine
XX residues.
XX
XX Example 25; Page 47; 102pp; English.
XX
XX The sequence represents a monomeric monocyclic peptide of the invention,
XX whose 3-dimensional structure is modelled on the expose loop of human
XX VEGF (vascular endothelial growth factor). The invention relates to a
XX method of producing a monomeric monocyclic peptide by a measuring beta-
XX beta carbon separation distances on opposite antiparallel strands of a
XX peptide loop fragment from an exposed loop of a growth factor protein and
XX cyclising the peptide by oxidising the cysteine residues. The monocyclic
XX peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic
XX peptides) and a cyclic peptide with at least one amino acid deleted prior
XX to cyclisation are used to interfere with angiogenesis,
XX neovascularisation or lymphangiogenesis in a mammal with a condition
XX characterised by angiogenesis, neovascularisation or lymphangiogenesis.
XX The condition is diabetic retinopathy, psoriasis, arthropathy,
XX hemangioma, vascularised malignant or benign tumour, post-recovery
XX cerebrovascular accident, post-angioplasty restenosis, head, heat or cold
XX trauma, substance-induced neovascularisation of the liver, excessive
XX hormone-related angiogenic dysfunction, diabetes induced neovascular
XX sequelae, hypertension induced neovascular sequelae, or chronic liver
XX infection. The peptides are also used to modulate vascular permeability
XX in a mammal (the mammal has a condition characterised by fluid
XX accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,
XX or brain. The peptides are used to image blood vessels and lymphatic
XX vasculature. The monomeric and bicyclic peptides are used to interfere
XX with at least one biological activity induced by VEGF, VEGF-C or -D and
XX are also used in combination with an anti-inflammatory agent, to treat a
XX chronic inflammation, especially rheumatoid arthritis, psoriasis and
XX diabetic retinopathy
XX
XX Sequence 11 AA;
XX
XX Query Match 74.1%; Score 41.5; DB 4; Length 11;
XX Best Local Similarity 72.7%; Pred. No. 2.5;
XX Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
XX
XX Qy 1 CISVPL-SVPC 10
XX |.|||||.|||
XX 1 CVSPLTTPC 11
XX
XX Db
XX
XX RESULT 7
XX AAU04544
XX ID AAU04544 standard; peptide; 11 AA.
XX
XX AC AAU04544;
XX
XX DT 26-SEP-2001 (first entry)
XX
XX DE VEGF based monocyclic peptide 22.
XX
XX Human; VEGF; vascular endothelial growth factor; angiogenesis;
XX neovascularisation; lymphangiogenesis; psoriasis; tumour;
XX diabetes induced neovascular sequelae; rheumatoid arthritis;

```



KW diabetic retinopathy; chronic inflammation; cyclic.

OS Synthetic.

XX Key Location/Qualifiers

FT Disulfide-bond 1..11

XX /note= "this bond cyclises the peptide"

XX WO200152875-A1.

XX 26-JUL-2001.

XX 18-JAN-2001; 2001WO-US001533.

XX 18-JAN-2000; 2000US-0176293P.

XX 16-MAY-2000; 2000US-0204590P.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Achen MG, Hughes RA, Stacker S, Cendron A;

XX WPI; 2001-442248/47.

XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis, or lymphangiogenesis, is produced by cyclizing a peptide loop fragment from an exposed loop of a growth factor protein by oxidizing the cysteine residues.

XX Example 25; Page 47; 102pp; English.

XX The sequence represents a monomeric monocyclic peptide of the invention, whose 3-dimensional structure is modelled on the expose loop of human VEGF (vascular endothelial growth factor). The invention relates to a method of producing a monomeric monocyclic peptide by a measuring beta-carbon separation distances on opposite antiparallel strands of a peptide loop fragment from an exposed loop of a growth factor protein and cyclising the peptide by oxidising the cysteine residues. The monocyclic peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic peptides) and a cyclic peptide with at least one amino acid deleted prior to cyclisation are used to interfere with angiogenesis, neovascularisation or lymphangiogenesis in a mammal with a condition characterised by angiogenesis, neovascularisation or lymphangiogenesis. The condition is diabetic retinopathy, psoriasis, arthropathy, hemangioma, vascularised malignant or benign tumour, post-recovery cerebrovascular accident, post-angioplasty restenosis, head, heat or cold trauma, substance-induced neovascularisation of the liver, excessive hormone-related angiogenic dysfunction, diabetes induced neovascular sequelae, hypertension induced neovascular sequelae, or chronic liver infection. The peptides are also used to modulate vascular permeability in a mammal (the mammal has a condition characterised by fluid accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura, or brain. The peptides are used to image blood vessels and lymphatic vasculature. The monomeric and bicyclic peptides are used to interfere with at least one biological activity induced by VEGF, VEGF-C or -D and are also used in combination with an anti-inflammatory agent, to treat a chronic inflammation, especially rheumatoid arthritis, psoriasis and diabetic retinopathy

XX Sequence 11 AA;

Query Match 72.3%; Score 40.5; DB 4; Length 11;  
Best Local Similarity 72.7%; Pred. No. 3.7;  
Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 1 CISVPL-SVPC 10

Db 1 CISLPISSVPC 11

RESULT 8

AAW95534

ID AAW95534 standard; protein; 77 AA.

XX

AC AAW95534;  
XX  
XX 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen SEQ ID NO: 4192.  
DE Human; reproductive system related antigen; reproductive system disorder;  
XX Human; gene therapy.  
KW  
XX  
OS Homo sapiens.  
XX  
XX WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001339.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189674P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.

PR	14-SEP-2000;	2000US-02334010P
PR	14-SEP-2000;	2000US-023340363P
PR	14-SEP-2000;	2000US-023340364P
PR	14-SEP-2000;	2000US-023340365P
PR	21-SEP-2000;	2000US-0234223P
PR	21-SEP-2000;	2000US-02342274P
PR	25-SEP-2000;	2000US-0234597P
PR	25-SEP-2000;	2000US-0234598P
PR	26-SEP-2000;	2000US-0234584P
PR	27-SEP-2000;	2000US-0235834P
PR	27-SEP-2000;	2000US-0235836P
PR	29-SEP-2000;	2000US-0236327P
PR	29-SEP-2000;	2000US-0236367P
PR	29-SEP-2000;	2000US-0236368P
PR	29-SEP-2000;	2000US-0236369P
PR	29-SEP-2000;	2000US-0236370P
PR	02-OCT-2000;	2000US-0236802P
PR	02-OCT-2000;	2000US-0237037P
PR	02-OCT-2000;	2000US-0237038P
PR	02-OCT-2000;	2000US-0237039P
PR	02-OCT-2000;	2000US-0237040P
PR	13-OCT-2000;	2000US-0239935P
PR	13-OCT-2000;	2000US-0239937P
PR	20-OCT-2000;	2000US-0241808P
PR	20-OCT-2000;	2000US-0241809P
PR	20-OCT-2000;	2000US-0241826P
PR	01-NOV-2000;	2000US-0244617P
PR	01-NOV-2000;	2000US-0244617P
PR	08-NOV-2000;	2000US-0246475P
PR	08-NOV-2000;	2000US-0246476P
PR	08-NOV-2000;	2000US-0246477P
PR	08-NOV-2000;	2000US-0246478P
PR	08-NOV-2000;	2000US-0246523P
PR	08-NOV-2000;	2000US-0246524P
PR	08-NOV-2000;	2000US-0246525P
PR	08-NOV-2000;	2000US-0246526P
PR	08-NOV-2000;	2000US-0246527P
PR	08-NOV-2000;	2000US-0246528P
PR	08-NOV-2000;	2000US-0246532P
PR	08-NOV-2000;	2000US-0246560P
PR	08-NOV-2000;	2000US-0246610P
PR	08-NOV-2000;	2000US-0246611P
PR	08-NOV-2000;	2000US-0246613P
PR	17-NOV-2000;	2000US-0249207P
PR	17-NOV-2000;	2000US-0249208P
PR	17-NOV-2000;	2000US-0249215P
PR	17-NOV-2000;	2000US-0249216P
PR	17-NOV-2000;	2000US-0249217P
PR	17-NOV-2000;	2000US-0249218P
PR	17-NOV-2000;	2000US-0249219P
PR	17-NOV-2000;	2000US-0249213P
PR	17-NOV-2000;	2000US-0249214P
PR	17-NOV-2000;	2000US-0249215P
PR	17-NOV-2000;	2000US-0249265P
PR	17-NOV-2000;	2000US-0249279P
PR	17-NOV-2000;	2000US-0249299P
PR	17-NOV-2000;	2000US-0249300P
PR	01-DEC-2000;	2000US-0250160P
PR	01-DEC-2000;	2000US-0250391P
PR	03-DEC-2000;	2000US-0251030P
PR	05-DEC-2000;	2000US-0251988P
PR	06-DEC-2000;	2000US-0256719P
PR	06-DEC-2000;	2000US-0251479P

```

08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254037P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-465570/50.
XX N-PSDB; AAL01504.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 11; SEQ ID NO 4192; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention
XX
XX Sequence 77 AA;
XX
XX Query Match 69.6%; Score 39; DB 4; Length 77;
XX Best Local Similarity 40.0%; Pred.No. 48;
XX Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 C1SVPLSVPC 10
XX |:|:|:|:|
XX DB 47 CLNAPIRIPC 56
XX
XX RESULT 9
XX ABB96216
XX ID ABB96216 standard; protein; 77 AA.
XX
XX AC ABB96216;
XX
XX DT 21-JUN-2002 (first entry)
XX
XX DE Human testicular antigen SEQ ID NO: 1600.
XX
XX KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic.
XX
XX OS Homo sapiens.
XX
XX FN WO200155317-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US001329.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209457P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.

```



```
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 C1SVPLSVPC 10
Db 47 CLNAPIRIPC 56

RESULT 10
AAM95827
ID AAM95827 standard; protein; 108 AA.
XX AC AAM95827;
XX DT 21-NOV-2001 (first entry)
XX DE Human reproductive system related antigen SEQ ID NO: 4485.
XX KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
OS Homo sapiens.
XX PN W0200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001339.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
```

```

PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX N-PSDB; AAL01797.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Claim 11; SEQ ID NO 4485; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a protein of the invention
XX
XX SQ Sequence 108 AA;
XX
XX Query Match 69.6%; Score 39; DB 4; Length 108;
XX Best Local Similarity 77.8%; Pred. No. 68;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 2 ISVPLSVPC 10
Db :|||||
33 VSVPLSHPC 41
XX
RESULT 11
ID ABB96358
ID ABB96358 standard; protein; 108 AA.
XX
XX ABB96358;
XX
XX AC
XX
XX DT 21-JUN-2002 (first entry)
XX
XX DE Human testicular antigen SEQ ID NO: 1742.
XX
XX KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic.
XX
XX OS Homo sapiens.
XX
XX PN WO20015317-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US001329.
XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209457P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216847P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.

```



CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 SQ Sequence 11 AA;

Query Match 68.8%; Score 38.5; DB 4; Length 11;  
 Best Local Similarity 63.6%; Pred. No. 7.8;  
 Matches 7; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

QY 1 CTSVPL-SVPC 10  
 ||::|| |::||  
 Db 1 CITIPLTSLPC 11

RESULT 13  
 ABG12542  
 ID ABG12542 standard; protein; 59 AA.

AC ABG12542;  
 XX  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #12533.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.

XX WO200175067-A2.  
 XX  
 XX 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 XX  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS76729.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

PS Claim 20; SEQ ID NO 42901; 103pp; English.  
 XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG0377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 59 AA;

Query Match 67.9%; Score 38; DB 4; Length 59;  
 Best Local Similarity 70.0%; Pred. No. 53;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTSVPLSVPC 10  
 ||::|| |::||  
 Db 9 CTSVLVPVPC 18

RESULT 14  
 ABG12530  
 ID ABG12530 standard; protein; 59 AA.

AC ABG12530;  
 XX  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #12521.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.

XX WO200175067-A2.  
 XX  
 XX 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 XX  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS76717.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

PS Claim 20; SEQ ID NO 42889; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 59 AA;

Query Match 67.9%; Score 38; DB 4; Length 59;  
Best Local Similarity 70.0%; Pred. No. 53;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 10  
||| : |||  
Db 9 CISVLVFPVC 18

RESULT 15

AA016003  
ID AA016003 standard; protein; 71 AA.

AC AA016003;

XX 12-OCT-2001 (first entry)

XX Peptide #2437 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000670.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human cervical epithelial cells.

XX Claim 27; SEQ ID NO 20829; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
XX (SENPs; see AAI10068-AA128459). The present sequence is a peptide encoded  
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
XX can be used to produce a single exon microarray, which can be used for  
XX measuring human gene expression in a sample derived from human cervical  
XX epithelial cells. By measuring gene expression, the probes are therefore  
XX useful in grading and/or staging of diseases of the cervix, notably  
XX cervical cancer. Note: The sequence data for this patent did not form  
XX part of the printed specification, but was obtained in electronic format  
XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 71 AA;

Query Match 67.9%; Score 38; DB 4; Length 71;  
Best Local Similarity 77.8%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 9  
||| : |||  
Db 51 CISVPLEPTP 59

Search completed: September 5, 2004, 09:55:21  
Job time : 34.8283 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 10:00:15 ; Search time 28.0808 Seconds  
(without alignments)  
112.199 Million cell updates/sec

Title: US-09-761-636A-13  
Perfect score: 56  
Sequence: 1 CISVPLSVPC 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	56	100.0	10	9	US-09-761-636A-13
2	45.5	81.2	11	9	US-09-761-636A-7
3	43.5	77.7	11	9	US-09-761-636A-23
4	43	76.8	140	16	US-10-437-963-174203
5	41.5	74.1	9	9	US-09-761-636A-14
6	41.5	74.1	11	9	US-09-761-636A-26
7	41	73.2	84	12	US-10-424-599-264847
8	40.5	72.3	11	9	US-09-761-636A-25
9	40	71.4	279	16	US-10-437-963-167987
10	39	69.6	77	10	US-09-764-891-4192
11	39	69.6	83	12	US-10-425-114-58210
12	39	69.6	108	10	US-09-764-891-4485
13	39	69.6	190	12	US-10-425-114-50298
14	38.5	68.8	11	9	US-09-761-636A-24
15	38	67.9	71	9	US-09-864-761-35705

16	38	67.9	142	12	US-10-424-599-220123	Sequence 220123,
17	38	67.9	202	16	US-10-437-963-140110	Sequence 140110,
18	38	67.9	217	15	US-10-104-047-2238	Sequence 2238, Ap
19	38	67.9	465	12	US-10-335-977-7910	Sequence 7910, Ap
20	38	67.9	578	12	US-10-335-977-7911	Sequence 7911, Ap
21	38	67.9	616	12	US-10-335-977-7912	Sequence 7912, Ap
22	38	67.9	1590	16	US-10-437-963-106960	Sequence 106960,
23	37	66.1	134	15	US-10-108-260A-4531	Sequence 4531, Ap
24	37	66.1	355	16	US-10-437-963-125857	Sequence 125857,
25	37	66.1	420	16	US-10-437-963-142446	Sequence 142446,
26	37	66.1	769	13	US-10-087-110-6	Sequence 6, Appli
27	37	66.1	875	13	US-10-087-110-2	Sequence 2, Appli
28	36	64.3	46	12	US-10-424-599-284943	Sequence 284943,
29	36	64.3	59	16	US-10-437-963-199587	Sequence 199587,
30	36	64.3	64	9	US-09-867-550-754	Sequence 754, App
31	36	64.3	70	14	US-10-029-386-29825	Sequence 29825, A
32	36	64.3	76	9	US-09-731-872-268	Sequence 268, App
33	36	64.3	82	16	US-10-437-963-106647	Sequence 106647,
34	36	64.3	108	12	US-10-424-599-184465	Sequence 184465,
35	36	64.3	113	9	US-09-815-248-2	Sequence 2, Appli
36	36	64.3	217	12	US-10-424-599-198582	Sequence 198582,
37	36	64.3	230	15	US-10-104-047-3554	Sequence 3554, Ap
38	36	64.3	241	12	US-10-425-114-60080	Sequence 60080, A
39	36	64.3	381	15	US-10-369-493-7846	Sequence 7846, Ap
40	36	64.3	669	10	US-09-790-852-7	Sequence 7, Appli
41	36	64.3	836	16	US-10-437-963-119123	Sequence 119123,
42	36	64.3	80	12	US-10-206-915-576	Sequence 576, App
43	35.5	63.4	80	12	US-10-199-670-576	Sequence 576, App
44	35.5	63.4	80	12	US-10-201-858-576	Sequence 576, App
45	35.5	63.4	80	12	US-10-201-858-576	Sequence 576, App

ALIGNMENTS

RESULT 1  
US-09-761-636A-13  
; Sequence 13, Application US/09761636A  
; Patent No. US20020085218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: GENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-761-636A-13

Query Match 100.0% Score 56; DB 9; Length 10;  
Best Local Similarity 100.0%; Pred.No. 0.023;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CISVPLSVPC 10  
| | | | | | | | | |  
Db 1 CISVPLSVPC 10

RESULT 2  
US-09-761-636A-7  
; Sequence 7, Application US/09761636A  
; Patent No. US20020085218A1

```
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-761-636A-7
```

```
Query Match 81.2%; Score 45.5; DB 9; Length 11;
Best Local Similarity 90.9%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
QY 1 C1SVPL-SVPC 10
| | | | | | | |
DB 1 C1SVPLTSVPC 11
```

```
RESULT 3
US-09-761-636A-23
; Sequence 23, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 11
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-761-636A-23
```

```
Query Match 77.7%; Score 43.5; DB 9; Length 11;
Best Local Similarity 81.8%; Pred. No. 2.3;
Matches 9; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
```

```
QY 1 C1SVPL-SVPC 10
| | | | | | | |
DB 1 C1SVPLTSVPC 11
```

```
RESULT 4
US-10-437-963-174203
; Sequence 174203, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
```

```
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174203
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: unsure
; LOCATION: (1)..(140)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72167C.1.pap
US-10-437-963-174203
```

```
Query Match 76.8%; Score 43; DB 16; Length 140;
Best Local Similarity 70.0%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 C1SVPLSVPC 10
| | | | | | | |
DB 115 CSSTPLTVPC 124
```

```
RESULT 5
US-09-761-636A-14
; Sequence 14, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-761-636A-14
```

```
Query Match 74.1%; Score 41.5; DB 9; Length 9;
Best Local Similarity 90.0%; Pred. No. 1.2e+06;
Matches 9; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
QY 1 C1SVPLSVPC 10
| | | | | | | |
DB 1 C1SVPL-VPC 9
```

```
RESULT 6
US-09-761-636A-26
; Sequence 26, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
```

; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09/761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-761-636A-26

Query Match 74.1%; Score 41.5; DB 9; Length 11;  
Best Local Similarity 72.7%; Pred. No. 4.7;  
Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 1 C1SVPL-SVPC 10  
Db 1 C1SVPLTTPC 11

## RESULT 7

US-10-424-599-264847  
; Sequence 264847, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 264847  
; LENGTH: 84  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_81177C.1.pap  
US-10-424-599-264847

Query Match 73.2%; Score 41; DB 12; Length 84;  
Best Local Similarity 70.0%; Pred. No. 39;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10  
Db 24 C1SVPLTTPC 33

## RESULT 8

US-09-761-636A-25  
; Sequence 25, Application US/09761636A  
; Patent No. US20020065218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09/761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18

; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 25  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-761-636A-25

Query Match 72.3%; Score 40.5; DB 9; Length 11;  
Best Local Similarity 72.7%; Pred. No. 6.7;  
Matches 8; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 1 C1SVPL-SVPC 10  
Db 1 C1SVPLSVPC 11

## RESULT 9

US-10-437-963-167987  
; Sequence 167987, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 167987  
; LENGTH: 279  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(279)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66546C.1.pap  
US-10-437-963-167987

Query Match 71.4%; Score 40; DB 16; Length 279;  
Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 9  
Db 175 C1SVPLSVPC 183

## RESULT 10

US-09-764-891-4192  
; Sequence 4192, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; PRIOR APPLICATION data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4192

```
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (38)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4192

Query Match          69.6%; Score 39; DB 10; Length 77;
Best Local Similarity 40.0%; Pred. No. 74;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLSVPC 10
Db 47 CLNAPIRPC 56

RESULT 11
US-10-425-114-58210
; Sequence 58210, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58210
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB84-029-E3_FLI.pep
US-10-425-114-58210

Query Match          69.6%; Score 39; DB 12; Length 83;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10
Db 68 TVPLSIPC 75

RESULT 12
US-09-764-891-4485
; Sequence 4485, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4485
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: SITE
; LOCATION: (49)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (51)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (73)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (93)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4485

Query Match          69.6%; Score 39; DB 10; Length 108;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISVPLSVPC 10
Db 33 VSVPLSHPC 41

RESULT 13
US-10-425-114-50298
; Sequence 50298, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 50298
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-012-E9_FLI.pep
US-10-425-114-50298

Query Match          69.6%; Score 39; DB 12; Length 190;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10
Db 175 TVPLSIPC 182

RESULT 14
US-09-761-636A-24
; Sequence 24, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
```

```
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
;   LENGTH: 11
;   TYPE: PRT
;   ORGANISM: synthetic construct
US-09-761-636A-24
```

```
Query Match      68.8%; Score 38.5; DB 9; Length 11;
Best Local Similarity 63.6%; Pred. No. 14;
Matches 7; Conservative 3; Mismatches 0; Indels 1; Gaps 1;
```

```
QY      1 CTSVPL-svpc 10
        ||:|||||:|
Db       1 CTTPLTSLPC 11
```

```
RESULT 15
US-09-864-761-35705
; Sequence 35705, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35705
; LENGTH: 71
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL078621.19
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EST HUMAN HIT: BF125501.1, EVALUE 1.70e+00
; OTHER INFORMATION: SWISSPROT HIT: P11992, EVALUE 2.20e-01
US-09-864-761-35705
```

```
Query Match      67.9%; Score 38; DB 9; Length 71;
Best Local Similarity 77.8%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 CTSVPLSVP 9
        |||||
Db       51 CTSVPLPTP 59
```

```
Search completed: September 5, 2004, 10:29:23
Job time : 29.0808 secs
```

**This Page Blank (uspto)**



```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23169

Query Match      69.6%; Score 39; DB 4; Length 306;
Best Local Similarity 70.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 C1SVPLSVPC 10
Db 7 CWSVPSAPC 16

RESULT 3
US-09-252-991A-23382
; Sequence 23382, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23382
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23382

Query Match      66.1%; Score 37; DB 4; Length 142;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 C1SVPLSVPC 10
Db 89 CLPPFISAPC 98

RESULT 4
US-09-252-991A-18804
; Sequence 18804, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18804
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18804

Query Match      66.1%; Score 37; DB 4; Length 154;
Best Local Similarity 50.0%; Pred. No. 57;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLSVPC 10
Db 137 CSSLPMLRPLC 146
```

```
RESULT 5
US-08-331-383-33
; Sequence 33, Application US/08331383
; Patent No. 5605809
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; TITLE OF INVENTION: Compositions for the Detection of
; TITLE OF INVENTION: Proteases in Biological Samples and Methods and Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,383
; FILING DATE: 28-OCT-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 16865-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-331-383-33

Query Match      64.3%; Score 36; DB 1; Length 9;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 3 SVPLSVPC 10
Db 2 AIPWSIPC 9

RESULT 6
US-08-549-008-43
; Sequence 43, Application US/08549008
; Patent No. 5714342
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of
; TITLE OF INVENTION: Protease in Biological Samples and Methods of Use Thereof
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/549,008  
FILING DATE: 27-OCT-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/331,383  
FILING DATE: 28-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen Lauver  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 016865-000110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-549-008-43

Query Match 64.3%; Score 36; DB 1; Length 9;  
Best Local Similarity 50.0%; Pred. No. 3e+05;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10  
Db 2 AIPMSIPC 9

RESULT 7  
US-08-802-981-143  
Sequence 143, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-802-981-143

Query Match 64.3%; Score 36; DB 3; Length 9;  
Best Local Similarity 50.0%; Pred. No. 3e+05;  
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10  
Db 2 AIPMSIPC 9

RESULT 8  
US-09-252-991A-29933  
Sequence 29933, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 29933  
LENGTH: 92  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29933

Query Match 64.3%; Score 36; DB 4; Length 92;  
Best Local Similarity 55.6%; Pred. No. 48;  
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISVPLSVPC 10  
Db 33 VEVPAVPC 41

RESULT 9  
US-09-621-976-5605  
Sequence 5605, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 5605  
LENGTH: 192  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: -25...-1  
US-09-621-976-5605

Query Match 64.3%; Score 36; DB 4; Length 192;  
Best Local Similarity 50.0%; Pred. No. 1e+02;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLSVPC 10  
Db 2 CWSLSMKVPC 11

```
RESULT 10
US-09-252-991A-29740
; Sequence 29740, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29740
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29740

Query Match      64.3%; Score 36; DB 4; Length 341;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      1 C1SVPLSVPC 10
      ||||| :|||
Db      294 C1SVSSTLPC 303

RESULT 11
US-09-252-991A-26663
; Sequence 26663, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26663
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26663

Query Match      64.3%; Score 36; DB 4; Length 510;
Best Local Similarity 50.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 C1SVPLSVPC 10
      ||||| :|||
Db      359 C1AFAAIPC 368

RESULT 12
US-09-252-991A-19384
; Sequence 19384, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
```

```
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19384
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19384

Query Match      64.3%; Score 36; DB 4; Length 700;
Best Local Similarity 75.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 SVPLSVPC 10
      : |||||
Db      46 AAPLSVPC 53

RESULT 13
US-09-252-991A-30161
; Sequence 30161, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30161
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30161

Query Match      62.5%; Score 35; DB 4; Length 183;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 C1SVPLSVPC 10
      ||||| :|||
Db      37 C1SRGMSVPC 46

RESULT 14
US-09-198-452A-145
; Sequence 145, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 145
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-145

Query Match      62.5%; Score 35; DB 4; Length 201;
```

Best Local Similarity 62.5%; Pred. No. 1.6e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 SVPLSVPC 10  
|:|:|  
Db 104 SIPVDVPC 111

QY 1 CIS-VPLSVPC 10  
|:|:|  
Db 7 CISTYPRSIIPC 17

Search completed: September 5, 2004, 10:22:01  
Job time : 9.88889 secs

RESULT 15  
US-08-722-719-58  
; Sequence 58, Application US/08722719  
; Patent No. 6001606  
; GENERAL INFORMATION:  
; APPLICANT: ROSEN, CRAIG A.  
; APPLICANT: RUBIN, STEVEN M.  
; APPLICANT: LI, HAO DONG  
; APPLICANT: ADAMS, MARK D.  
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS FOR  
; TREATING DISEASE STATES WITH MYELOID PROGENITOR INHIBITORY  
; TITLE OF INVENTION: FACTOR-1 (MIPF-1), MONOCYTE COLONY INHIBITORY FACTOR  
; TITLE OF INVENTION: (M-CIF), AND MACROPHAGE INHIBITORY FACTOR-4 (MIP-4)  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/722,719  
; FILING DATE: 30-SEP-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,209  
; FILING DATE: 22-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/208,339  
; FILING DATE: 08-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/446,881  
; FILING DATE: 05-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/465,682  
; FILING DATE: 06-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/468,775  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: STEFFE, ERIC K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0330007  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 72 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-722-719-58

Query Match 61.6%; Score 34.5; DB 3; Length 72;  
Best Local Similarity 63.6%; Pred. No. 65;  
Matches 7; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:47:29 ; Search time 6.90909 Seconds  
(without alignments)  
125.302 Million cell updates/sec

Title: US-09-761-636A-14  
Perfect score: 52  
Sequence: 1 CISVPLVPC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	75.0	297	2 T25817	hypothetical prote
2	39	75.0	347	2 A81988	probable N-acetyl-
3	39	75.0	347	2 D81043	N-acetyl-gamma-glu
4	39	75.0	416	2 A61404	keratin A, type I
5	39	75.0	416	2 A46559	keratin, type I, h
6	38	73.1	154	2 D82795	phage-related endo
7	38	73.1	577	2 T00416	hypothetical prote
8	38	73.1	1015	2 C84918	probable Arp-depen
9	38	73.1	3623	2 T08618	intrinsic factor-B
10	37	71.2	287	2 H64690	type IIS restricti
11	37	71.2	407	2 T19895	hypothetical prote
12	36	69.2	1639	2 T50119	probable sensory t
13	35	67.3	61	2 JQ1086	nodulin-14 precus
14	35	67.3	278	2 AF0936	probable Arac-fam
15	35	67.3	305	2 A82728	conserved hypothet
16	35	67.3	496	2 S51668	tyrosine kinase -
17	35	67.3	1210	2 D88013	protein K10B4.1 [i
18	35	67.3	1440	1 SYHUQT	multifunctional am
19	35	67.3	2240	2 T37057	probable multi-dom
20	34	65.4	397	2 AC1877	hypothetical prote
21	34	65.4	551	2 JC7562	glioblastoma RING
22	34	65.4	1876	2 T13801	phosphoinositide 3
23	33	63.5	339	2 A13609	daunorubicin resis
24	33	63.5	370	2 T25627	hypothetical prote
25	33	63.5	434	2 T01390	hypothetical prote
26	33	63.5	491	2 T46915	hypothetical prote
27	33	63.5	497	1 S53834	NADH2 dehydrogenas
28	33	63.5	498	2 T45465	acetate CoA-transf
29	33	63.5	725	2 JE0099	neural cell adhesi

30	33	63.5	771	2 F70757	probable cation tr
31	33	63.5	956	2 A65072	hypothetical prote
32	33	63.5	956	2 B91098	probable dehydroge
33	33	63.5	956	2 F85943	probable dehydroge
34	33	63.5	957	2 T41246	probable ribonucle
35	33	63.5	1021	2 S26985	probable DNA-direc
36	33	63.5	1046	2 A26838	prestalk protein p
37	33	63.5	1253	1 VHWV	structural polypro
38	33	63.5	1827	2 A35694	cuti protein - fis
39	33	63.5	1828	2 T41455	cuti protein - fis
40	32	61.5	58	2 A23101	lon protein - Esch
41	32	61.5	72	1 W5WL39	E5 protein - human
42	32	61.5	84	2 A61594	thrombospondin - b
43	32	61.5	115	2 D71541	probable inclusion
44	32	61.5	174	2 T02683	hypothetical prote
45	32	61.5	185	1 A47196	dual specificity p

## ALIGNMENTS

## RESULT 1

T25817  
hypothetical protein K12D9.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25817

R:Graves, T.

submitted to the EMBL Data Library, November 1996

A:Description: The sequence of C. elegans cosmid K12D9.

A:Reference number: Z20093

A:Accession: T25817

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-297 <GRA>

A:Cross-references: EMBL:U80030; PIDN:AAB37600.1; GSPDB:GN00023; CESP:K12D9.3

A:Experimental source: strain Bristol N2; clone K12D9

C:Genetics:

A:Gene: CESP:K12D9.3

A:Map position: 5

A:Introns: 165/3; 208/3

Query Match 75.0%; Score 39; DB 2; Length 297;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9

Db 169 CISLPIVTC 177

## RESULT 2

A81988  
probable N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) NMA0676 [imported]

C:Species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: A81988

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A:Reference number: A81775; MUID:20222556; PMID:10761919

A:Accession: A81988

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-347 <PAR>

A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83963.1; PID:g737940

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: argC; NMA0676

C:Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase

C:Keywords: oxidoreductase

Query Match 75.0%; Score 39; DB 2; Length 347;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVP 8  
 |:|:|:|  
 Db 156 CVSPLVP 163

## RESULT 3

D81043  
 N-acetyl-gamma-glutamyl-phosphate reductase NMB1787 [imported] - Neisseria meningitidis  
 C;Species: Neisseria meningitidis  
 C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
 C;Accession: D81043  
 R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.  
 Science 287, 1809-1815, 2000  
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A;Reference number: A81000; MUID:20175755; PMID:10710307  
 A;Accession: D81043  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-347 <TET>  
 A;Cross-references: GB:AE002528; GB:AE002098; NID:g7227034; PIDN:AAF42126.1; PID:g722704  
 A;Experimental source: serogroup B, strain MC58  
 C;Genetics:  
 A;Gene: NMB1787  
 C;Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase

Query Match 75.0%; Score 39; DB 2; Length 347;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVP 8  
 |:|:|:|  
 Db 156 CVSPLVP 163

## RESULT 4

A61404  
 keratin A, type I - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 09-Sep-1994 #sequence\_revision 09-Sep-1994 #text\_change 03-May-1996  
 C;Accession: A61404  
 R;Kaytes, P.S.; McNab, A.R.; Rea, T.J.; Groppi, V.; Kawabe, T.T.; Buhl, A.E.; Bertolino,  
 J. Invest. Dermatol. 97, 835-842, 1991  
 A;Title: Hair-specific keratins: characterization and expression of a mouse type I kerat  
 A;Reference number: A61404; MUID:92013235; PMID:1717610  
 A;Accession: A61404  
 A;Status: preliminary; not compared with conceptual translation  
 A;Molecule type: DNA  
 A;Residues: 1-416 <KAY>  
 C;Superfamily: cytoskeletal keratin

Query Match 75.0%; Score 39; DB 2; Length 416;  
 Best Local Similarity 66.7%; Pred. No. 19;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLVP 9  
 |:|:|:|  
 Db 385 CVSPLVP 393

## RESULT 5

A46559  
 keratin, type I, hair (clone MHKA-1) - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 13-Aug-1999  
 C;Accession: A46559  
 R;Bertolino, A.P.; Checkla, D.M.; Notterman, R.; Sklaver, I.; Schiiff, T.A.; Freedberg, I

J. Invest. Dermatol. 91, 541-546, 1988  
 A;Title: Cloning and characterization of a mouse type I hair keratin cDNA.  
 A;Reference number: A46559; MUID:89054860; PMID:2461417  
 A;Accession: A46559  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-416 <BER>  
 A;Cross-references: GB:M27734; NID:g198585; PIDN:AAA39372.1; PID:g198586  
 C;Superfamily: cytoskeletal keratin

Query Match 75.0%; Score 39; DB 2; Length 416;  
 Best Local Similarity 66.7%; Pred. No. 19;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLVP 9  
 |:|:|:|  
 Db 385 CVSPLVP 393

## RESULT 6

D82795  
 Phage-related endolysin XF0513 [imported] - Xylella fastidiosa (strain 9a5c)  
 C;Species: Xylella fastidiosa  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 23-Mar-2001  
 C;Accession: D82795  
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A;Reference number: A82515; MUID:20365717; PMID:10910347  
 A;Note: for a complete list of authors see reference number A59328 below  
 A;Accession: D82795  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-154 <SIM>  
 A;Cross-references: GB:AE003900; GB:AE003849; NID:g9105366; PIDN:AAF83323.1; GSPDB:GN001  
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Carrer, F.  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrato, D.M.; Carrer, F.  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000

C;Superfamily: phage T4 lysozyme; phage T4 lysozyme homology

Query Match 73.1%; Score 38; DB 2; Length 154;  
 Best Local Similarity 75.0%; Pred. No. 11;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVP 8  
 |:|:|:|  
 Db 78 CISVPLVP 85

## RESULT 7

T00416  
 hypothetical protein T30B22.1 - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 24-Nov-1999  
 C;Accession: T00416  
 R;Rounsley, S.D.; Iain, X.; Ketchum, K.A.; Crooby, M.L.; Brandon, R.C.; Sykes, S.M.; Mas  
 submitted to the EMBL Data Library, October 1999  
 A;Description: Arabidopsis thaliana chromosome II BAC T30B22 genomic sequence.  
 A;Reference number: Z14149

A:Accession: T00416  
 A:Status: translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-577 <ROU>  
 A:Cross-references: EMBL:AC002535; NID:g2529657; PID:g2529658  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 2  
 A:Introns: 87/3; 126/3; 175/1; 264/3; 329/3; 507/3  
 C:Superfamily: Arabidopsis thaliana hypothetical protein T30B22.1

Query Match 73.1%; Score 38; DB 2; Length 577;  
 Best Local Similarity 75.0%; Pred. No. 40;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 C1SVPLVP 8  
 |:|||||  
 Db 238 C1SVPLVP 245

RESULT 8  
 C84918  
 probable ATP-dependent RNA helicase A [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84918  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10611797  
 A:Accession: C84918  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1015 <STO>  
 A:Cross-references: GB:AE002093; NID:g3738282; PIDN:AAC63624.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g47680  
 A:Map position: 2

Query Match 73.1%; Score 38; DB 2; Length 1015;  
 Best Local Similarity 75.0%; Pred. No. 68;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 C1SVPLVP 8  
 |:|||||  
 Db 676 C1SVPLVP 683

RESULT 9  
 T08618  
 intrinsic factor-B12 receptor CUBILIN precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 02-Aug-2002  
 C:Accession: T08618  
 R:Moestrup, S.K.; Kozvriki, R.; Kristiansen, M.; Kayesen, J.H.; Rasmussen, H.H.; Brault,  
 J. Biol. Chem. 273, 5235-5242, 1998  
 A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibodies  
 A:Reference number: Z16459; MUID:98148073; PMID:9478979  
 A:Accession: T08618  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3623 <MOE>  
 A:Cross-references: EMBL:AF022247; NID:g3834379; PIDN:AAC71661.1; PID:g383434380  
 C:Genetics:  
 A:Gene: CUBILIN  
 C:Superfamily: intrinsic factor-B12 receptor cubilin; EGF homology  
 C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membrane  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-3623/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <MAT>  
 F:133-164/Domain: EGF homology <EGF1>  
 F:436-467/Domain: EGF homology <EGF>

Query Match 73.1%; Score 38; DB 2; Length 3623;  
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 C1SVPLVPC 9  
 |:|||||  
 Db 313 C1SVPLVPC 321

RESULT 10  
 H64690  
 type IIS restriction enzyme M2 protein - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 22-Jun-2003  
 C:Accession: H64690  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;  
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne  
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.  
 Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: H64690  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-287 <TOM>  
 A:Cross-references: GB:AE000637; GB:AE000511; NID:g2314536; PIDN:AAD08412.1; PID:g231453  
 C:Genetics:  
 A:Start codon: TTG  
 C:Superfamily: type II site-specific DNA-methyltransferase

Query Match 71.2%; Score 37; DB 2; Length 287;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 C1SVPLVP 8  
 |:|||||  
 Db 86 C1SVPLVP 93

RESULT 11  
 T19895  
 hypothetical protein C41G7.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T19895  
 R:Steward, C.  
 submitted to the EMBL Data Library, October 1996

A:Reference number: Z19192  
 A:Accession: T19895  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-407 <WIL>  
 A:Cross-references: EMBL:Z81048; PIDN:CAB02841.1; GSPDB:GN00019; CESP:C41G7.4  
 A:Experimental source: clone C41G7  
 C:Genetics:  
 A:Gene: CESP:C41G7.4  
 A:Map position: 1  
 A:Introns: 27/2; 107/3; 151/3; 333/3; 373/3

Query Match 71.2%; Score 37; DB 2; Length 407;  
 Best Local Similarity 66.7%; Pred. No. 43;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLVPC 9  
 |:|||||  
 Db 113 C1SVPLVPC 121

RESULT 12  
 T50119

probable sensory transduction histidine kinase [imported] - fission yeast (Schizosacchar

C;Species: Schizosaccharomyces pombe  
C;Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C;Accession: T50119  
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, February 2000  
A;Reference number: Z25039  
A;Accession: T50119  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1639 <SEE>  
A;Cross-references: EMBL:AL157734; PIDN: CAB75776.1; GSPDB: GN00066; SPDB: SPAC1834.08  
A;Experimental source: strain 972h(-); cosmid c1834  
C;Genetics:  
A;Gene: SPDB: SPAC1834.08  
A;Map position: 1  
A;Introns: 23/3

Query Match 69.2%; Score 36; DB 2; Length 1639;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLVPC 9  
Db 557 SVPLVPC 563  
|||||

## RESULT 13

QJ1086  
nodulin-14 precursor - garden pea (fragment)  
C;Species: Pisum sativum (garden pea)  
C;Date: 23-Nov-1991 #sequence\_revision 23-Nov-1991 #text\_change 20-Apr-2001  
C;Accession: QJ1086  
R;Scheres, B.; van Engelen, F.; van der Knaap, E.; van de Wiel, C.; van Kammen, A.; Biss  
Plant Cell 2, 687-700, 1990  
A;Title: Sequential induction of nodulin gene expression in the developing pea nodule.  
A;Reference number: QJ1084; MUID: 93005665; PMID: 2152123  
A;Accession: QJ1086  
A;Molecule type: mRNA  
A;Residues: 1-61 <SCH>  
A;Cross-references: GB: S45162; NID: G257036; PIDN: AAB23538.1; PID: G257037  
A;Experimental source: root nodule  
C;Comment: The gene encoding this protein is expressed in the second stage of root nodul  
C;Genetics:  
A;Gene: ENOD14  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-61/Product: nodulin 14 #status predicted <MAT>

Query Match 67.3%; Score 35; DB 2; Length 61;  
Best Local Similarity 71.4%; Pred. No. 16;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLVPC 9  
Db 21 NIPLVPC 27  
:|||||

## RESULT 14

AF0936  
Probable AraC-family transcription regulator STY3756 [imported] - Salmonella enterica su  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 08-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AF0936  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moutle, S.; O'Gea, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID: 21534947; PMID: 11677608  
A;Accession: AF0936  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-278 <PAR>  
A;Cross-references: GB: AL513382; PIDN: CAD09512.1; PID: g16504629; GSPDB: GN00176  
C;Genetics:  
A;Gene: STY3756  
C;Superfamily: hypothetical protein b2382

Query Match 67.3%; Score 35; DB 2; Length 278;  
Best Local Similarity 62.5%; Pred. No. 68;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ISVPLVPC 9  
Db 56 ITAPLIPC 63  
|||:

## RESULT 15

AB2728  
conserved hypothetical protein XFI068 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C;Accession: AB2728  
R;anonymous, The xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: AB2515; MUID: 20365717; PMID: 10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: AB2728  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-305 <SIM>  
A;Cross-references: GB: AE003943; GB: AE003849; NID: g9106006; PIDN: AAF83878.1; GSPDB: GN001  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; F  
briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, F  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;  
A;Reference number: A59328  
A;Contents: annotation  
C;Genetics:  
A;Gene: XFI068

Query Match 67.3%; Score 35; DB 2; Length 305;  
Best Local Similarity 57.1%; Pred. No. 74;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLVPC 9  
Db 3 SIBIIPC 9  
|||:

Search completed: September 5, 2004, 10:01:25  
Job time : 7.90909 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:38:39 ; Search time 4.09091 Seconds  
(without alignments)  
114.554 Million cell updates/sec

Title: US-09-761-636A-14  
Perfect score: 52  
Sequence: 1 CISVPLVPC 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	39	75.0	347	1	ARGC_NEIMA
2	39	75.0	347	1	ARGC_NEIMB
3	39	75.0	416	1	KIM1_MOUSE
4	35	67.3	61	1	NO14_PEA
5	35	67.3	78	1	VE5_HPV70
6	35	67.3	279	1	MOVE_CMYAS
7	35	67.3	490	1	FA10_RABIT
8	35	67.3	551	1	OCN3_RAT
9	35	67.3	556	1	OCN3_HUMAN
10	35	67.3	1108	1	EMBA_MYCAV
11	35	67.3	1440	1	SYEP_HUMAN
12	34	65.4	551	1	RN27_HUMAN
13	34	65.4	551	1	RN27_MOUSE
14	34	65.4	715	1	BBS2_BRARE
15	34	65.4	966	1	M172_HUMAN
16	34	65.4	1173	1	TSP1_XENLA
17	33	63.5	497	1	NU4M_ACACA
18	33	63.5	771	1	CTPG_MYCTU
19	33	63.5	956	1	XDHD_ECO57
20	33	63.5	956	1	XDHD_ECOLI
21	33	63.5	1021	1	DPOM_NEUCR
22	33	63.5	1046	1	PSTA_DICDI
23	33	63.5	1253	1	POLS_SFV
24	33	63.5	1828	1	CUT1_SCHPO
25	32	61.5	72	1	VE5_HPV39
26	32	61.5	174	1	TP14_ARATH
27	32	61.5	183	1	AROK_SYNEL
28	32	61.5	185	1	DUS3_HUMAN
29	32	61.5	185	1	DUS3_MOUSE
30	32	61.5	282	1	PLC3_CABEL
31	32	61.5	305	1	CORA_HPBEDU
32	32	61.5	312	1	NRT1_CHICK
33	32	61.5	346	1	GATD_ECOLI

34	32	61.5	387	1	DHB2_HUMAN	P37059	homo sapien
35	32	61.5	413	1	GAT1_MOUSE	P17679	mus musculus
36	32	61.5	413	1	GAT1_RAT	P43429	rattus norv
37	32	61.5	459	1	RSP6_CHLRE	Q01657	chlamydomon
38	32	61.5	467	1	KIM6_HUMAN	O76013	homo sapien
39	32	61.5	649	1	ACSA_VIBCH	Q9KX59	vibrio chol
40	32	61.5	662	1	T9S2_MOUSE	P58021	mus musculus
41	32	61.5	663	1	T9S2_HUMAN	Q99805	homo sapien
42	32	61.5	824	1	DD20_HUMAN	Q9UHI6	homo sapien
43	32	61.5	1170	1	TSP1_BOVIN	Q28178	bos taurus
44	32	61.5	1170	1	TSP1_HUMAN	P07996	homo sapien
45	32	61.5	1170	1	TSP1_MOUSE	P35441	mus musculus

## ALIGNMENTS

RESULT 1  
ARGC\_NEIMA  
ID \_ARGC\_NEIMA STANDARD; PRT; 347 AA.  
AC Q9UV06;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-  
DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).  
DE ARGC OR NMA0676.  
GN Neisseria meningitidis (serogroup A).  
OS Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;  
RX MEDLINE=20222556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,  
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrell B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
meningitidis Z2491".  
RL Nature 404:502-506(2000).  
CC -|- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+) + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.  
CC -|- PATHWAY: Arginine biosynthesis; third step.  
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -|- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily 1.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: AL162753; CAB83963.1; -  
PIR: A81988; A81988.  
HAMAP: MF\_00150; -, 1.  
InterPro: IPR000706; AGPR act\_site.  
InterPro: IPR000534; Semialdh dh.  
Pfam: PF01118; Semialdehyde dh; 1.  
Pfam: PF02774; Semialdehyde\_dhc; 1.  
ProDom: PD003765; AGPR\_act\_site; 1.  
PROSITE: PS01224; AGRC; FALSE NEG.  
KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.  
FT ACT SITE 152 152 BY SIMILARITY.  
SQ SEQUENCE 347 AA; 37227 MW; E11E8A19A22B9246 CRC64;

Query Match 75.0%; Score 39; DB 1; Length 347;

Best Local Similarity 75.0%; Pred. No. 4.2;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 C1SVPLVP 8  
|:|:|  
Db 156 CVSPLPVP 163

## RESULT 2

ARGC NEIMB STANDARD; PRT; 347 AA.  
AC 09JY18;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).  
GN ARGC OR NM1787.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / Serogroup B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwin M.L., DeBoo R., Peterson J.D., Hickey E.K., Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B., Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.; RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";  
RL Science 287:1809-1815(2000).  
CC -!- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+) + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.  
CC -!- PATHWAY: Arginine biosynthesis; third step.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily 1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE002528; AAF42126.1; -.  
CC PIR; D81043; D81043.  
CC TIGR; NMB1787; -.  
CC HAMAP; MF 00150; -; 1.  
CC InterPro; IPR000706; AGPR act site.  
CC InterPro; IPR000534; Semialdh\_ch.  
CC Pfam; PF01118; Semialdehyde\_dh; 1.  
CC Pfam; PF02774; Semialdehyde\_dhC; 1.  
CC ProDom; PD003765; AGPR act site; 1.  
CC PROSITE; PS01224; AGRC; FALSE NEG.  
KW Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.  
FT ACT\_SITE 152 152 BY SIMILARITY.  
FT SITE 152 152  
SQ SEQUENCE 347 AA; 37213 MW; A927C570DE19ECD1 CRC64;

Query Match 75.0%; Score 39; DB 1; Length 347;  
Best Local Similarity 75.0%; Pred. No. 4.2;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 C1SVPLVP 8  
|:|:|  
Db 156 CVSPLPVP 163

## RESULT 3

KIMI\_MOUSE STANDARD; PRT; 416 AA.  
ID KIMI\_MOUSE  
AC Q61765;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Keratin, type I cuticular HA1 (Hair keratin, type I HA1) (HKA-1).  
GN KRT1A1 OR KRT1-1 OR HKAL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Hair;  
RX MEDLINE=89054860; PubMed=2461417;  
RA Bertolino A.P., Checkla D.M., Notterman R., Sklaver I., Schiff T.A., Freedberg I.M., Didona G.J.;  
RT "Cloning and characterization of a mouse type I hair keratin cDNA.";  
RL J. Invest. Dermatol. 91:541-546(1988).  
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF HAIR/MICROFIBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC).  
CC -!- SIMILARITY: Belongs to the intermediate filament family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M27734; AAA39372.1; -.  
CC PIR; A46559; A46559.  
CC MGD; MGI:130993; Krt1-1.  
CC InterPro; IPR001664; IF.  
CC InterPro; IPR002957; Keratin\_1.  
CC Pfam; PF00038; filament; 1.  
CC PRINTS; PR01248; TYPE1KERATIN.  
CC PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Keratin.  
FT DOMAIN 2 56  
FT DOMAIN 57 363 ROD.  
FT DOMAIN 364 416 TAIL.  
FT DOMAIN 57 91 COIL 1A.  
FT DOMAIN 92 102 LINKER 1.  
FT DOMAIN 103 203 COIL 1B.  
FT DOMAIN 204 219 LINKER 12.  
FT DOMAIN 220 363 COIL 2.  
FT SITE 305 305 STUTTER.  
SQ SEQUENCE 416 AA; 47112 MW; 9C3759CF4E3A3EF3 CRC64;

Query Match 75.0%; Score 39; DB 1; Length 416;  
Best Local Similarity 66.7%; Pred. No. 5.1;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 C1SVPLVPC 9  
|:|:|  
Db 385 CVSNPCVPC 393

## RESULT 4

NOL14\_PEA STANDARD; PRT; 61 AA.  
ID NOL14\_PEA  
AC P26415;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Nodulin 14 precursor (N-14) (Fragment).  
GN ENOD14.

```
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Sparkle; TISSUE=Root nodules;
RX MEDLINE=93005665; PubMed=21512123;
RA Scheres B., van Engelen F., van der Knaap E., van de Wiel C.,
RA van Kammen A., Bisseling T.;
RT "Sequential induction of nodulin gene expression in the developing
RT pea nodule.";
RL Plant Cell 2:687-700(1990).
CC -1- DEVELOPMENTAL STAGE: Expressed in the second stage of root nodule
CC formation.
CC -1- MISCELLANEOUS: CONTAINS 4 CYSTEINES ARRANGED IN TWO PAIRS IN SUCH
CC A WAY THAT THEY MIGHT BE CAPABLE OF BINDING A METAL ION.
CC -1- SIMILARITY: TO NODULIN 3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S45162; AAB23538.1; -.
DR PIR; JQ1086; JQ1086.
KW Nodulation; Signal; Metal-binding.
FT NON_TER 1 1
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 61 NODULIN 14.
SQ SEQUENCE 61 AA; 7175 MW; D508059695F62B8B CRC64;

Query Match 67.3%; Score 35; DB 1; Length 61;
Best Local Similarity 71.4%; Pred. No. 4;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SVPLVPC 9
:|||||
Db 21 NIDLVPC 27

RESULT 5
VE5 HPV70
ID VE5 HPV70 STANDARD; PRT; 78 AA.
AC P50774;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable E5 protein.
GN E5.
OS Human papillomavirus type 70.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=39457;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96249586; PubMed=8815087;
RA Forslund O., Hansson B.G.;
RT "Human papillomavirus type 70 genome cloned from overlapping PCR
RT products: complete nucleotide sequence and genomic organization.";
RL J. Clin. Microbiol. 34:802-809(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
CC -----
DR EMBL; U01941; AAC54855.1; -.
DR InterPro; IPR004270; Papilloma_E5.
DR Pfam; PF03025; Papilloma_E5; 1.
KW Early protein.
SQ SEQUENCE 78 AA; 9262 MW; DFEA59CB355042D1 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 78;
Best Local Similarity 75.0%; Pred. No. 5.2;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CISVPLVP 8
:|||||
Db 24 CCSVPLLP 31

RESULT 6
MOVLP CMVAS
ID MOVLP CMVAS STANDARD; PRT; 279 AA.
AC O36639;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cell-to-cell movement protein (MP) (3A protein).
GN 3A.
OS Cucurbit mosaic virus (strain As) (CMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Cucumovirus.
OX NCBI_TaxID=117118;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim S.H., Park Y.I.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
CC INFECTED CELLS TO ADJACENT CELLS.
CC -1- SIMILARITY: Belongs to the cucumoviruses/bromoviruses 3A family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP013291; AAB67166.1; -.
DR InterPro; IPR000603; 3A_mov.
DR Pfam; PF00803; 3A; 1.
KW DNA-binding; Transport.
SQ SEQUENCE 279 AA; 30509 MW; 2B47059288DC1069 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 279;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9
:|||||
Db 85 CVSPQVLC 93

RESULT 7
FA10 RABIT
ID FA10 RABIT STANDARD; PRT; 490 AA.
AC O19045;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Coagulation factor X precursor (EC 3.4.21.6) (Stuart factor).
GN F10.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
```

RN SEQUENCE FROM N.A. PubMed-9101642;  
 RX MEDLINE=97256311; Anderson K.D., James H.L.;  
 RA Chendurthi U.R., Anderson K.D., James H.L.;  
 RT "Characterization of a full-length cDNA for rabbit factor X.";  
 RL Thromb. Res. 85:503-514(1997).  
 CC -!- FUNCTION: Factor Xa is a vitamin K-dependent glycoprotein that  
 CC converts prothrombin to thrombin in the presence of factor Va,  
 CC calcium and phospholipid during blood clotting.  
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Thr and then  
 CC Arg-|-Ile bonds in prothrombin to form thrombin.  
 CC -!- SUBUNIT: The two chains are formed from a single-chain precursor  
 CC by the excision of two Arg residues and are held together by 1 or  
 CC more disulfide bonds.  
 CC -!- PTM: The vitamin K-dependent, enzymatic carboxylation of some  
 CC glutamate residues allows the modified protein to bind calcium (By  
 CC similarity).  
 CC -!- PTM: N- and O-glycosylated (By similarity).  
 CC -!- PTM: THE ACTIVATION PEPTIDE IS CLEAVED BY FACTOR IXA (IN THE  
 CC INTRINSIC PATHWAY), OR BY FACTOR VIIA (IN THE EXTRINSIC PATHWAY)  
 CC (BY SIMILARITY).  
 CC -!- MISCELLANEOUS: Calcium also binds, with stronger affinity to  
 CC another site, beyond the GLA domain.  
 CC -!- SIMILARITY: Belongs to peptidase family S1.  
 CC -!- SIMILARITY: Contains 2 EGF-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>  
 CC or send an email to [license@iesb-sib.ch](mailto:license@iesb-sib.ch)).  
 CC -----  
 DR EMBL; AF003200; AAB62542.1; -;  
 DR HSSP; P00742; IHCG.  
 DR MEROPS; S01.216; -;  
 DR InterPro; IPR000152; Asx hydroxyl S.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF II.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR002383; GLA blood.  
 DR InterPro; IPR001254; Peptidase S1.  
 DR InterPro; IPR001314; Peptidase\_SIA.  
 DR InterPro; IPR000294; VitK\_dep\_GLA.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF00594; gla; 1.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR00010; EGFBLD.  
 DR PRINTS; PR00001; GLABLOOD.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00069; GLA; 1.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS00026; EGF\_2; 2.  
 DR PROSITE; PS01186; EGF\_3; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS00011; GLU\_CARBOXYLATION; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Glycoprotein; Hydrolase; Serine protease; Plasma; Blood coagulation;  
 KW Gamma-carboxyglutamic acid; Hydroxylation; Calcium-binding; Vitamin K;  
 KW Signal; Zymogen; EGF-like domain; Repeat.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT PROPEP 21 40 BY SIMILARITY.  
 FT CHAIN 41 180 FACTOR X LIGHT CHAIN.  
 FT CHAIN 184 490 FACTOR X HEAVY CHAIN.  
 FT PROPEP 184 232 ACTIVATION PEPTIDE.

FT CHAIN	233	490	ACTIVATED FACTOR XA, HEAVY CHAIN.
FT DOMAIN	86	122	EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN	125	165	EGF-LIKE 2.
FT MOD_RES	233	490	SERINE PROTEASE.
FT MOD_RES	46	46	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	47	47	SIMILARITY).
FT MOD_RES	47	47	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	54	54	SIMILARITY).
FT MOD_RES	56	56	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	59	59	SIMILARITY).
FT MOD_RES	60	60	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	65	65	SIMILARITY).
FT MOD_RES	66	66	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	69	69	SIMILARITY).
FT MOD_RES	72	72	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	75	75	SIMILARITY).
FT MOD_RES	79	79	GAMMA-CARBOXYGLUTAMIC ACID (BY
FT MOD_RES	103	103	SIMILARITY).
FT ACT_SITE	274	274	HYDROXYLATION (BY SIMILARITY).
FT ACT_SITE	320	320	CHARGE RELAY SYSTEM.
FT ACT_SITE	417	417	CHARGE RELAY SYSTEM.
FT DISULFID	90	101	BY SIMILARITY.
FT DISULFID	95	110	BY SIMILARITY.
FT DISULFID	112	121	BY SIMILARITY.
FT DISULFID	129	140	BY SIMILARITY.
FT DISULFID	136	149	BY SIMILARITY.
FT DISULFID	151	164	BY SIMILARITY.
FT DISULFID	172	340	INTERCHAIN (BY SIMILARITY).
FT DISULFID	239	244	BY SIMILARITY.
FT DISULFID	259	275	BY SIMILARITY.
FT DISULFID	388	402	BY SIMILARITY.
FT DISULFID	413	441	BY SIMILARITY.
FT CARBOHYD	61	61	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	205	205	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE	490 AA; 53965 MW; 3A39FA85AF2A6D11 CRC64;		

Query Match 67.3%; Score 35; DB 1; Length 490;  
 Best Local Similarity 66.7%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 CISVPLVPC 9  
 Db 164 CISTELEPFC 172  
 RESULT 8  
 OCN3 RAT STANDARD; PRT; 551 AA.  
 ID OCN3 RAT  
 AC O89446;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Organic cation transporter 3 (Solute carrier family 22, member 3).  
 GN SLC22A3 OR OCT3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 NCBI\_TaxID=10116;  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;

RX MEDLINE=98298098; PubMed=9632645;  
 RA Kekuda R., Prasad P.D., Wu X., Wang H., Fei Y.-J., Leibach F.H.,  
 RA Ganapathy V.;  
 RT "Cloning and functional characterization of a potential-sensitive,  
 RT polyspecific organic cation transporter (OCT3) most abundantly  
 RT expressed in placenta.";  
 RL J. Biol. Chem. 273:15971-15979(1998).  
 RN [2]  
 RP FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99047659; PubMed=9830022;  
 RA Wu X., Kekuda R., Huang W., Fei Y.-J., Leibach F.H., Chen J.,  
 RA Conway S.-J., Ganapathy V.;  
 RT "Identity of the organic cation transporter OCT3 as the extraneuronal  
 RT monoamine transporter (uptake2) and evidence for the expression of  
 RT the transporter in the brain";  
 RL J. Biol. Chem. 273:32776-32786(1998).  
 RN [2]  
 RP FUNCTION: Mediates potential-dependent transport of a variety of  
 CC organic cations. May play a significant role in the disposition of  
 CC cationic neurotoxins and neurotransmitters in the brain.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PLACENTA WITH MODERATE  
 CC EXPRESSION IN INTESTINE, HEART AND BRAIN. IN THE BRAIN, WIDELY  
 CC EXPRESSED, ESPECIALLY IN HIPPOCAMPUS, CEREBELLUM AND CEREBRAL  
 CC CORTEX. EXPRESSION IS LOW IN KIDNEY AND LUNG AND UNDETECTABLE IN  
 CC LIVER.  
 CC -1- SIMILARITY: Belongs to the sugar transporter family. Organic  
 CC cation subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF055286; AAC40150.1; -;  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR004749; Orgcat\_transp.  
 DR InterPro; IPR005828; Sub\_transporter.  
 DR InterPro; IPR005829; Sug\_transporter.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 DR TIGRFAMs; TIGR00898; 2A0119; 1.  
 DR PROSITE; PS50850; MFS; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE NEG.  
 KM Transport; Ion transport; Transmembrane; Glycoprotein.  
 FT TRANSMEM 21 41  
 FT TRANSMEM 177 197 POTENTIAL.  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 376 396 POTENTIAL.  
 FT TRANSMEM 463 483 POTENTIAL.  
 FT TRANSMEM 493 513 POTENTIAL.  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 551 AA; 61049 MW; 23F0BC24E4F3DE1 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 551;

Best Local Similarity 85.7%; Pred.No. 39;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 SVPLVPC 9

Db 116 SAPLVPC 122

RESULT 9

OCCN3\_HUMAN

ID OCN3\_HUMAN STANDARD; PRT; 556 AA.  
 AC 075751; QSUP02;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Organic cation transporter 3 (Extraneuronal monoamine transporter)  
 DE (EMT) (Solute carrier family 22, member 3).  
 GN SLC22A3 OR EMT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=Kidney;  
 RX MEDLINE=99212254; PubMed=10196521;  
 RA Gruendemann D., Schechinger B., Rappold G.A., Schoemig E.,  
 RA "Molecular identification of the corticosterone-sensitive  
 RA extraneuronal catecholamine transporter.";  
 RL Nat. Neurosci. 1:349-351(1998).  
 RN [2]  
 RP SEQUENCE OF 513-556 FROM N.A., AND TISSUE SPECIFICITY.  
 RX MEDLINE=99134299; PubMed=9933568;  
 RA Verhaagh S., Schweifer N., Barlow D.P., Zwart R.;  
 RA "Cloning of the mouse and human solute carrier 22a3 (SLC22a3/SLC22A3)  
 RA identifies a conserved cluster of three organic cation transporters on  
 RA mouse chromosome 17 and human 6q26-q27.";  
 RL Genomics 55:209-218(1999).  
 RN [3]  
 RP FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=Kidney;  
 RX MEDLINE=20425388; PubMed=10966924;  
 RA Wu X., Huang W., Ganapathy M.E., Wang H., Kekuda R., Conway S.-J.,  
 RA Leibach F.H., Ganapathy V.;  
 RT "Structure, function, and regional distribution of the organic cation  
 RT transporter OCT3 in the kidney.";  
 RL Am. J. Physiol. 279:F449-F458(2000).  
 CC -1- FUNCTION: Mediates potential-dependent transport of a variety of  
 CC organic cations. May play a significant role in the disposition of  
 CC cationic neurotoxins and neurotransmitters in the brain.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, SKELETAL MUSCLE,  
 CC PROSTATE, AORTA, LIVER, FETAL LUNG, SALIVARY GLAND, ADRENAL GLAND,  
 CC KIDNEY AND BRAIN CORTEX. NO EXPRESSION DETECTED IN SPLEEN.  
 CC -1- SIMILARITY: Belongs to the sugar transporter family. Organic  
 CC cation subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ001417; CAA04751.1; -;  
 DR EMBL; AF078749; AAD20977.1; -;  
 DR Genew; HGNC:10967; SLC22A3.  
 DR MIM; 604842; -;  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0005624; C:membrane fraction; TAS.  
 DR GO; GO:0015101; F:organic cation transporter activity; TAS.  
 DR GO; GO:0015695; P:organic cation transport; TAS.  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR004749; Orgcat\_transp.  
 DR InterPro; IPR005828; Sub\_transporter.  
 DR InterPro; IPR005829; Sug\_transporter.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 DR TIGRFAMs; TIGR00898; 2A0119; 1.  
 DR PROSITE; PS50850; MFS; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE NEG.  
 KM Transport; Ion transport; Transmembrane; Glycoprotein.

```
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 468 488 POTENTIAL.
FT TRANSMEM 498 518 POTENTIAL.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 556 AA; 61279 MW; C3CA2D77DD21C658 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 556;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 SVPLVPC 9
Db 121 SAPLVPC 127

RESULT 10
EMBA MYCAV STANDARD; PRT; 1108 AA.
ID EMBA MYCAV STANDARD; PRT; 1108 AA.
AC P71485;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable arabinosyltransferase A (EC 2.4.2.-).
DN EMBA
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2151;
RX MEDLINE=97030297; PubMed=8876238;
RA Belanger A.E., Besra G.S., Ford M.E., Mikusova K., Belisle J.T.,
RA Brennan P.J., Inamine J.M.;
RT "The embA genes of Mycobacterium avium encode an arabinosyl
RT transferase involved in cell wall arabinan biosynthesis that is the
RT target for the antimycobacterial drug ethambutol."
RL Proc. Natl. Acad. Sci. U.S.A. 93:11919-11924(1996).
CC -!- FUNCTION: Arabinosyl transferase responsible for the
CC polymerization of arabinose into the arabinan of arabinogalactan.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE EMB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U66560; AAC44547.1; -.
DR InterPro; IPR007680; Arab transf.
DR Pfam; PF04602; arab transf; 1.
KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
KW Antibiotic resistance.
FT TRANSMEM 12 34 POTENTIAL.
FT TRANSMEM 204 223 POTENTIAL.
FT TRANSMEM 258 280 POTENTIAL.
FT TRANSMEM 334 356 POTENTIAL.
FT TRANSMEM 368 387 POTENTIAL.
FT TRANSMEM 397 414 POTENTIAL.
FT TRANSMEM 421 443 POTENTIAL.
FT TRANSMEM 463 482 POTENTIAL.
FT TRANSMEM 531 553 POTENTIAL.
FT TRANSMEM 582 604 POTENTIAL.
```

```
FT TRANSMEM 616 638 POTENTIAL.
FT TRANSMEM 653 675 POTENTIAL.
FT TRANSMEM 696 718 POTENTIAL.
SQ SEQUENCE 1108 AA; 117404 MW; 0ED8E648768075FA CRC64;

Query Match 67.3%; Score 35; DB 1; Length 1108;
Best Local Similarity 62.5%; Pred. No. 81;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CISVPLVP 8
Db 28 CLAVPLLP 35

RESULT 11
SYEP HUMAN STANDARD; PRT; 1440 AA.
ID SYEP HUMAN STANDARD; PRT; 1440 AA.
AC P07814;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bifunctional aminoacyl-tRNA synthetase [Includes: Glutamyl-tRNA
DE synthetase (EC 6.1.1.17) (Glutamate-tRNA ligase); Prolyl-tRNA
DE synthetase (EC 6.1.1.15) (Proline-tRNA ligase)].
DN EPRS OR QPRS OR GLNS OR PARS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91107633; PubMed=1988429;
RA Fett R., Knippers R.;
RT "The primary structure of human glutamyl-tRNA synthetase. A highly
RT conserved core, amino acid repeat regions, and homologies with
RT translation elongation factors."
RL J. Biol. Chem. 266:1448-1455(1991).
RN [2]
RP PRELIMINARY SEQUENCE OF 96-887 FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=88262551; PubMed=3290852;
RA Thoenes P., Fett R., Schray B., Kunze N., Knippers R.;
RT "The core region of human glutamyl-tRNA synthetase homologies with
RT the Escherichia coli and yeast enzymes."
RL Nucleic Acids Res. 16:5391-5406(1988).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) = AMP +
CC diphosphate + L-prolyl-tRNA(Pro).
CC -!- SUBUNIT: Component of the multisynthetase complex which is
CC comprised of a bifunctional glutamyl-prolyl-tRNA synthetase, the
CC monospecific isoleucyl-leucyl, glutamyl, methionyl, lysyl,
CC arginyl, and aspartyl-tRNA synthetases as well as three auxiliary
CC proteins, p18, p48 and p43.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -!- SIMILARITY: Contains 3 WHEP-TFS domains.
CC -!- CAUTION: Was originally thought to be a glutamyl-tRNA
CC synthetase.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a number of
CC sequence errors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X54326; CAA38224.1; -.
```

```

DR EMBL; X07466; CAA30354.1; ALT_SEQ.
DR PIR; A38663; SYHUQT.
DR PDB; 1FYJ; 31-DEC-02.
DR Genew; HGNC:3418; EPRS.
DR MIM; 138295; .
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005625; C:soluble fraction; TAS.
DR GO; GO:0006461; P:protein complex assembly; TAS.
DR InterPro; IPR004526; Glx arch.
DR InterPro; IPR000924; Glu ERNA-synt_1c.
DR InterPro; IPR004046; GST-cterm.
DR InterPro; IPR004154; HGP anticonodon.
DR InterPro; IPR004499; ProS fam I.
DR InterPro; IPR002314; tRNA-synt 2b.
DR InterPro; IPR001412; tRNA-synt I.
DR InterPro; IPR002316; tRNA-synt_pro.
DR InterPro; IPR006195; tRNA-synt_II.
DR InterPro; IPR000738; WHEP-TRS.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF03129; HGTP anticonodon; 1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c_C; 1.
DR Pfam; PF00587; tRNA-synt_2b; 1.
DR Pfam; PF00458; WHEP-TRS; 3.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR PRINTS; PR01046; TRNASYNTHPRO.
DR TIGRFAMS; TIGR00463; gltx arch; 1.
DR TIGRFAMS; TIGR00408; proS fam I; 1.
DR PROSITE; PS00178; AA-TRNA-LIGASE I; 1.
DR PROSITE; PS00862; AA-TRNA-LIGASE II; 1.
DR PROSITE; PS00762; WHEP-TRS; 3.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Multifunctional enzyme; Repeat; 3D-structure.
FT DOMAIN 92 687
FT DOMAIN 588 884
FT SITE 132 142
FT SITE 360 364
FT BINDING 363 363
FT DOMAIN 688 733
FT DOMAIN 761 806
FT DOMAIN 839 884
FT DOMAIN 887 919
SQ SEQUENCE 1440 AA; 163026 MW; C4E185A0AA41C204 CRC64;

Query Match 67.3%; Score 35; DB 1; Length 1440;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9
Db 1229 CVQWIIIPC 1237

RESULT 12
RN27_HUMAN
ID RN27_HUMAN STANDARD; PRT; 551 AA.
AC Q9BZE9; Q9C028;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE RING finger protein 27 (Globlastoma-expressed ring finger protein)
DE (Tripartite motif-containing protein 8).
GN TRIM8 OR RN27 OR GERP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=20568703; PubMed=11118312;
RA Vincent S.R., Kwasnicka D.A., Fretier P.;

```

```

RT "A novel RING finger-B box-coiled-coil protein, GERP.";
RL Biochem. Biophys. Res. Commun. 279:482-486(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuron;
RX MEDLINE=21231161; PubMed=11331580;
RA Raymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA Minucci S., Pellicci P.G., Ballabio A.;
RT "The tripartite motif family identifies cell compartments.";
RL EMBL J. 20:2140-2151(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heiton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Widely expressed.
CC -1- DOMAIN: The coiled coil domain is required for homodimerization.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -1- SIMILARITY: Contains 2 B box-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF281046; AAG53087.1; -.
DR EMBL; AF220034; AAG53488.1; -.
DR EMBL; BC021925; AAH21925.1; -.
DR PIR; JC7562; JC7562.
DR Genew; HGNC:15579; TRIM8.
DR MIM; 606125; -.
DR GO; GO:0008270; F:zinc ion binding; NAS.
DR InterPro; IPR000315; Znf_Bbox.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00643; zf-B_box_1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00119; ZF_BBOX; FALSE_NEG.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Nuclear protein; Coiled coil; Zinc-finger; Repeat.
FT ZN_FING 15 56
FT ZN_FING 92 132
FT ZN_FING 140 182
FT DOMAIN 181 249
FT CONFLICT 174 174 H -> R (IN REF. 1).
SQ SEQUENCE 551 AA; 61489 MW; 1FEF89029EB9BACB CRC64;

```

1.1 INTRODUCTION





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:46:09 ; Search time 20.3636 Seconds  
(without alignments)  
139.448 Million cell updates/sec

Title: US-09-761-636A-14  
Perfect score: 52  
Sequence: 1 CISVPLVPC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_protist.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	78.8	342	2 Q53028	Q53028 rhodococcus
2	40	76.9	648	10 Q9NA88	Q9NA88 arabidopsis
3	39	75.0	362	5 P91378	P91378 caenorhabdi
4	38	73.1	80	2 Q9Z1Y6	Q9Z1Y6 borrelia bu
5	38	73.1	134	4 Q8N719	Q8N719 homo sapien
6	38	73.1	154	16 Q9PF23	Q9PF23 xylella fas
7	38	73.1	198	16 Q30782	Q30782 chlamydophi
8	38	73.1	562	5 Q9VN94	Q9VN94 drosophila
9	38	73.1	568	5 Q966C5	Q966C5 caenorhabdi
10	38	73.1	1015	10 Q22243	Q22243 arabidopsis
11	38	73.1	3623	11 Q70244	Q70244 rattus norv
12	37	71.2	132	4 Q8N7H7	Q8N7H7 homo sapien
13	37	71.2	283	4 Q86S05	Q86S05 homo sapien
14	37	71.2	287	11 Q9QY31	Q9QY31 mus musculu
15	37	71.2	287	11 Q8C244	Q8C244 mus musculu
16	37	71.2	287	16 Q25921	Q25921 helicobacte

17	37	71.2	407	5 Q93368	Q93368 caenorhabdi
18	37	71.2	972	10 Q7X9C0	Q7X9C0 lotus japon
19	36	69.2	114	12 Q96736	Q96736 fowl adenov
20	36	69.2	147	11 Q8BU71	Q8BU71 mus musculu
21	36	69.2	150	16 Q8R673	Q8R673 fusobacteri
22	36	69.2	185	16 Q9Y311	Q9Y311 streptococc
23	36	69.2	186	16 Q8NZF5	Q8NZF5 streptococc
24	36	69.2	309	16 Q87F37	Q87F37 xylella fas
25	36	69.2	442	11 Q8BN56	Q8BN56 mus musculu
26	36	69.2	556	5 Q15716	Q15716 dictyosteli
27	36	69.2	583	4 Q8N787	Q8N787 homo sapien
28	36	69.2	650	17 Q8TVZ0	Q8TVZ0 methanopyru
29	36	69.2	945	10 Q9M7A8	Q9M7A8 nicotiana t
30	36	69.2	1348	5 Q86KE8	Q86KE8 dictyosteli
31	36	69.2	1639	3 Q9P7Q7	Q9P7Q7 schizosacch
32	35	67.3	120	4 Q8NBV6	Q8NBV6 homo sapien
33	35	67.3	143	4 Q9NQC0	Q9NQC0 homo sapien
34	35	67.3	186	4 Q86SP2	Q86SP2 homo sapien
35	35	67.3	197	5 Q9BKR0	Q9BKR0 caenorhabdi
36	35	67.3	249	4 Q9BPY7	Q9BPY7 homo sapien
37	35	67.3	265	4 Q9BYT4	Q9BYT4 homo sapien
38	35	67.3	278	16 Q8Z306	Q8Z306 salmonella
39	35	67.3	283	16 Q8ZKM1	Q8ZKM1 salmonella
40	35	67.3	305	16 Q9PEG0	Q9PEG0 xylella fas
41	35	67.3	309	4 Q9BX88	Q9BX88 homo sapien
42	35	67.3	388	4 Q8TDU8	Q8TDU8 homo sapien
43	35	67.3	388	5 Q9UGV3	Q9UGV3 strongyloid
44	35	67.3	422	16 Q81PZ5	Q81PZ5 bacillus an
45	35	67.3	496	10 Q01863	Q01863 phytophthor

## ALIGNMENTS

RESULT 1

Q53028	PRELIMINARY;	PRT;	342 AA.
AC Q53028			
DT 01-NOV-1996	(TrEMBLrel. 01, Created)		
DT 01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DT 01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE Reductase.			
GN AMOD			
OS Rhodococcus corallinus.			
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC Corynebacterineae; Gordoniaceae; Gordonia.			
OX NCBI_TaxID=36822;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=B-276;			
RA Saeki H., Furuhashi K.;			
RT "Cloning and characterization of a Nocardia corallina B-276 gene cluster encoding alkene monooxygenase.";			
RL J. Ferment. Bioeng. 78:339-406(1994).;			
CC -!- COFACTOR: BINDS 1 2PE-2S CLUSTER (BY SIMILARITY).			
DR EMBL; D37875; BAA07115.1; -.			
DR HSSP; P23486; LQPU.			
DR GO; GO:0005489; F:electron transporter activity; IEA.			
DR GO; GO:0016491; F:oxidoreductase activity; IEA.			
DR GO; GO:0006118; P:electron transport; IEA.			
DR InterPro; IPR006058; 2Pe2S fd BS.			
DR InterPro; IPR001834; Cyt B5_reductase.			
DR InterPro; IPR008333; FAD_binding_6.			
DR InterPro; IPR001041; Ferredoxin.			
DR InterPro; IPR001709; FPN cvt reductase.			
DR InterPro; IPR001433; Oxred FAD/NAD(P).			
DR InterPro; IPR001221; Phe hydroxylase.			
DR Pfam; PF00970; FAD_binding_6; 1.			
DR Pfam; PF00111; fer2; 1.			
DR Pfam; PF00175; NAD binding 1; 1.			
DR PRINTS; PR00406; CYTB5RDTASE.			
DR PRINTS; PR00371; FPNCR.			
DR PRINTS; PR00410; PHEHYDRLASE.			

```

DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron; Iron-sulfur.
SQ SEQUENCE 342 AA; 37282 MW; 4651F38D116A8049 CRC64;

Query Match 78.8%; Score 41; DB 2; Length 342;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9
| | | | |
Db 75 CASVPLVPC 83

RESULT 2
Q9MA88 PRELIMINARY; PRT; 648 AA.
AC Q9MA88;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T12H1.28 protein.
GN T12H1.28.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv Columbia;
RA Lin X.; Kaul S.; Town C.D.; Benito M.; Creasy T.H.; Haas B.; Wu D.;
RA Rowning C.M.; Koo H.; Fujii C.Y.; Utterback T.R.; Barnstead M.E.;
RA Bowman C.L.; White O.; Niernan W.C.; Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T12H1 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009177; AAF27037.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 2.
DR PROSITE; PS00018; EF_HAND; 2.
SQ SEQUENCE 648 AA; 73066 MW; D22FD23B724D8B7B CRC64;

Query Match 76.9%; Score 40; DB 10; Length 648;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9
| | | | |
Db 227 CFTPLVPC 235

RESULT 3
P91378 PRELIMINARY; PRT; 362 AA.
ID P91378
AC P91378;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Serpentine receptor, class w protein 121.
GN SRW-121.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequencing of the nematode C. elegans: a platform for
RL Science 282:2012-2018 (1998).
RN [2]

```

---

```

RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Graves T.;
RT "The sequence of C. elegans cosmid K12D9.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U80030; AAG24161.2; -.
KW Receptor.
SQ SEQUENCE 362 AA; 41292 MW; F25CFD3B0E1D9435 CRC64;

Query Match 75.0%; Score 39; DB 5; Length 362;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9
| | | | |
Db 233 CISLPIVTC 241

RESULT 4
Q9Z1Y6 PRELIMINARY; PRT; 80 AA.
ID Q9Z1Y6
AC Q9Z1Y6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N40;
RT "G5 clone from Borrelia burgdorferi.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005056; AAD01230.1; -.
KW Hypothetical protein.
SQ SEQUENCE 80 AA; 9171 MW; C61A1E359DED8EA3 CRC64;

Query Match 73.1%; Score 38; DB 2; Length 80;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 8
| | | | |
Db 40 CISVPLVPC 47

RESULT 5
Q8N7I9 PRELIMINARY; PRT; 134 AA.
ID Q8N7I9
AC Q8N7I9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ25492.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;

```

RA Nishi T., Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,  
 RA Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,  
 RA Kawakami B., Nagai K., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK098358; BAC05291.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 134 AA; 13911 MW; 9605A9F9912C2802 CRC64;

Query Match 73.1%; Score 38; DB 4; Length 134;  
 Best Local Similarity 77.8%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9  
 |||||  
 Db 94 CISVTLSPC 102

RESULT 6  
 Q9PFZ3 PRELIMINARY; PRT; 154 AA.  
 AC Q9PFZ3;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Phage-related endolysin.  
 GN XF0513.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=95C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Buato M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Coutinho L.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AS003900; AAF83323.1; -;  
 DR F01; D82795;  
 DR GO; GO:0003796; F:lysozyme activity; IEA.  
 DR GO; GO:0016998; P:cell wall catabolism; IEA.  
 DR GO; GO:0009253; P:peptidoglycan catabolism; IEA.  
 DR InterPro; IPR002196; Glyco\_hydro\_24.  
 DR Pfam; PF00959; Phage\_lysozyme; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 154 AA; 16793 MW; 20A5570A19D715FD CRC64;

Query Match 73.1%; Score 38; DB 16; Length 154;  
 Best Local Similarity 75.0%; Pred. No. 19;  
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVP 8  
 |||||  
 Db 78 CISVPLMP 85

RESULT 7  
 O30782 PRELIMINARY; PRT; 198 AA.  
 AC O30782;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Inclusion membrane protein B.  
 GN INCB OR CCA00491.  
 OS Chlamydomophila caviae.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 OX NCBI\_TaxID=83557;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GPIC;  
 RX MEDLINE=98326820; PubMed=9663687;  
 RA Bannantine J.P., Rockey D.D., Hackstadt T.;  
 RT "Fandem genes of Chlamydia psittaci that encode proteins localized to  
 RT the inclusion membrane.";  
 RL Mol. Microbiol. 28:1017-1026(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GPIC;  
 RX MEDLINE=22569155; PubMed=12682364;  
 RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,  
 RA Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,  
 RA Unayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,  
 RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,  
 RA Fraser C.M.;  
 RT "Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC):  
 RT examining the role of niche-specific genes in the evolution of the  
 RT Chlamydiaceae.";  
 RL Nucleic Acids Res. 31:2134-2147(2003).  
 DR EMBL; AF017105; AAC46378.1; -;  
 DR EMBL; AB016995; AAP09235.1; -;  
 DR TIGR; CCA00491; -;  
 KW Complete proteome.  
 SQ SEQUENCE 198 AA; 20917 MW; 455E5EA09444FE29 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 198;  
 Best Local Similarity 55.8%; Pred. No. 25;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9  
 |||||  
 Db 157 CISLPIISC 165

RESULT 8  
 Q9VN94 PRELIMINARY; PRT; 562 AA.  
 AC Q9VN94;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE CG1113 protein.  
 GN CG1113.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galle R.P.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam A.R., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
RA Foaler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusakern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003603; AAF52054.1; -.  
DR FlyBase; FBgn037304; CG1113.  
SQ SEQUENCE 562 AA; 63406 MW; AC09F5B6F5D2F970 CRC64;

Query Match 73.1%; Score 38; DB 5; Length 562;  
Best Local Similarity 55.6%; Pred. NO. 68;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9  
Db 9 CVKVPIDPC 17

RESULT 9  
Q966C5 PRELIMINARY; PRT; 568 AA.  
AC Q966C5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein Y24D9A.2.  
GN Y24D9A.2  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Wilson R.;  
RT "Genome sequence of the nematode *C. elegans*: a platform for

RT investigating biology. The *C. elegans* Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Cordes M., Wohldmann P.;  
RT "The sequence of *C. elegans* cosmid Y24D9A.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006729; AAK84599.2; -.  
DR WormPep; Y24D9A.2; CE33849.  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS02080; SET; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 568 AA; 65908 MW; EE4299B1DCFB98D CRC64;

Query Match 73.1%; Score 38; DB 5; Length 568;  
Best Local Similarity 66.7%; Pred. No. 69;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9  
Db 277 CSSKPIVPC 285

RESULT 10  
Q22243 PRELIMINARY; PRT; 1015 AA.  
ID O22243  
AC O22243;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-AUG-1998 (TrEMBLrel. 11, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Putative ATP-dependent RNA helicase A.  
GN ATG24680.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083487; PubMed=10617197;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.";  
RL Nature 402:761-768(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC005309; AAC63624.1; -.  
DR PIR; C84918; C84918.  
DR PIR; T00416; T00416.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.

```

DR InterPro; IPR000571; Znf_CCH.
DR Pfam; PF00271; heliCase_G; 1.
DR Pfam; PF00642; zf-CCCH_2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00356; Znf_C3H1; 2.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 1015 AA; 115084 MW; D1C342B338C561C9 CRC64;

Query Match 73.1%; Score 38; DB 10; Length 1015;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 8
Db 676 CUSVFPVP 683

RESULT 11
O70244 PRELIMINARY; PRT; 3623 AA.
AC O70244;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Intrinsic factor-B12 receptor precursor.
GN CUBILIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=98148073; PubMed=9478979;
RA Moestrup S.K., Kozyraki R., Kristiansen M., Kayesen J.H.,
RA Rasmussen H.H., Brault D., Pontillon F., Goda F.O., Christensen E.I.,
RA Hammond T.G., Verroust P.J.;
RT "The intrinsic factor-vitamin B12 receptor and target of teratogenic
RT antibodies is a megalin-binding peripheral membrane protein with
RT homology to developmental proteins.";
RL J. Biol. Chem. 273:5235-5242(1998).
CC -!- SIMILARITY: CONTAINS 26 CUB DOMAINS.
DR EMBL; AF022247; AAC71661.1; -.
DR PIR; T08618; T08618.
DR HSP; P00740; IEDM.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00431; CUB; 27.
DR Pfam; PF00008; EGF; 5.
DR SMART; SM00042; CUB; 26.
DR SMART; SM00179; EGF_CA; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS01180; CUB; 27.
DR PROSITE; PS00022; EGF_1; 4.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 4.
KW EGF-like domain; Receptor; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3623 INTRINSIC FACTOR-B12 RECEPTOR.
SQ SEQUENCE 3623 AA; 398981 MW; 39FB792AC6545240 CRC64;

Query Match 73.1%; Score 38; DB 11; Length 3623;
Best Local Similarity 66.7%; Pred. No. 4.1e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9
Db 313 CSQAPLVPC 321

DR InterPro; IPR000571; Znf_CCH.
DR Pfam; PF00271; heliCase_G; 1.
DR Pfam; PF00642; zf-CCCH_2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00356; Znf_C3H1; 2.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 1015 AA; 115084 MW; D1C342B338C561C9 CRC64;

Query Match 73.1%; Score 37; DB 4; Length 283;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9
Db 313 CSQAPLVPC 321

```

```

RESULT 12
Q8N7H7 PRELIMINARY; PRT; 132 AA.
AC Q8N7H7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ25553.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RX TISSUE=Thyroid;
RA Nimomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098419; BAC05303.1; -.
KW Hypothetical protein.
SQ SEQUENCE 132 AA; 14716 MW; 8C2661CA0E87309B CRC64;

Query Match 71.2%; Score 37; DB 4; Length 132;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ISVPLVPC 9
Db 121 LAVPLIPC 128

```

```

RESULT 13
Q86SU5 PRELIMINARY; PRT; 283 AA.
AC Q86SU5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to snail homolog 3 (Drosophila) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RX TISSUE=Lung, Spleen, and Petal;
RA Trauberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC041461; AAH41461.1; -.
DR Genex; HGNC:18411; SNAI3.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2_5.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
FT NON_TER 1 1
SQ SEQUENCE 283 AA; 31413 MW; 5376EAE364E86373 CRC64;

Query Match 71.2%; Score 37; DB 4; Length 283;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ISVPLVPC 9
Db 121 LAVPLIPC 128

```

```
QY 1 CISVPLVP 8
Db 58 CISLPLLP 65

RESULT 14
Q9QY31 PRELIMINARY; PRT; 287 AA.
AC Q9QY31;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Smail-related zinc finger protein SMUC (Zinc finger protein 293).
GN SNAI3 OR ZFP293 OR SMUC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK089285; BAC40828.1; -.
DR MGD; MGI:1353563; Sna13.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003700; F:transcription factor activity; IDA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR InterPro; IPR000345; CyfC_heme_BS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
SQ SEQUENCE 287 AA; 31662 MW; B3E42A6FF9BAF1FE CRC64;

Query Match 71.2%; Score 37; DB 11; Length 287;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVP 8
Db 67 CISLPLLP 74

Search completed: September 5, 2004, 10:00:09
Job time : 22.3636 secs
```

```
QY 1 CISVPLVP 8
Db 58 CISLPLLP 65

RESULT 14
Q9QY31 PRELIMINARY; PRT; 287 AA.
AC Q9QY31;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Smail-related zinc finger protein SMUC (Zinc finger protein 293).
GN SNAI3 OR ZFP293 OR SMUC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6;
RX MEDLINE=20075155; PubMed=10606664;
RA Kataoka H., Murayama T., Yokode M., Mori S., Sano H., Ozaki H.,
RA Yokota Y., Nishikawa S.I., Kita T.;
RT "A novel Snail-related transcription factor Smuc regulates basic
RT helix-loop-helix transcription factor activities via specific B-box
RT motifs.";
RL Nucleic Acids Res. 28:626-633(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AF133714; AAF22956.1; -.
DR EMBL; AK042234; BAC31201.1; -.
DR HSSP; P15822; 1BEO.
DR MGD; MGI:1353563; Sna13.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0003700; F:transcription factor activity; IDA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 287 AA; 31636 MW; 02B42A6F8BAFD48 CRC64;

Query Match 71.2%; Score 37; DB 11; Length 287;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVP 8
Db 67 CISLPLLP 74

RESULT 15
Q8C244 PRELIMINARY; PRT; 287 AA.
AC Q8C244;
DT 01-MAY-2000 (TrEMBLrel. 23, Created)
DT 01-MAY-2000 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zinc finger protein 293.
GN SNAI3.
```





CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis,  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 XX Sequence 9 AA;

Query Match 100.0%; Score 52; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9  
 |||||  
 Db 1 CISVPLVPC 9

RESULT 2  
 AAU04532  
 ID AAU04532 standard; peptide; 10 AA.  
 XX  
 AC AAU04532;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE VEGF based monocyclic peptide 10.  
 XX  
 KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
 KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
 KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
 KW diabetic retinopathy; chronic inflammation; cyclic.  
 XX  
 OS Synthetic.

XX Key Location/Qualifiers  
 XX Disulfide-bond 1..10  
 FT /note= "This bond cyclises the peptide"  
 FT  
 XX

PN WO200152875-A1.  
 XX  
 PD 26-JUL-2001.  
 XX

PF 18-JAN-2001; 2001WO-US0001533.  
 XX  
 PR 18-JAN-2000; 2000US-0176293P.  
 PR 16-MAY-2000; 2000US-0204590P.  
 XX

XX (LUDW-) LUDWIG INST CANCER RES.  
 XX

XX Achen MG, Hughes RA, Stacker S, Cendron A;  
 PI  
 XX WPI; 2001-442248/47.  
 DR

XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
 PT or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
 PT from an exposed loop of a growth factor protein by oxidizing the cysteine

PT residues.

XX Claim 49; Page 32; 102pp; English.

XX The sequence represents a monomeric monocyclic peptide of the invention,  
 CC whose 3-dimensional structure is modelled on the expose loop of human  
 CC VEGFD (vascular endothelial growth factor). The invention relates to a  
 CC method of producing a monomeric monocyclic peptide by a measuring beta-  
 CC beta carbon separation distances on opposite antiparallel strands of a  
 CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis,  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX

XX Sequence 10 AA;

Query Match 79.8%; Score 41.5; DB 4; Length 10;  
 Best Local Similarity 90.0%; Pred. No. 2.2; 0; Indels 1; Gaps 1;  
 Matches 9; Conservative 0; Mismatches 0;

QY 1 CISVPL-VPC 9  
 |||||  
 Db 1 CISVPLSVPC 10

RESULT 3

AAU04526  
 ID AAU04526 standard; peptide; 11 AA.

XX AAU04526;

XX 26-SEP-2001 (first entry)

XX VEGF based monocyclic peptide 3.

XX Human; VEGF; vascular endothelial growth factor; angiogenesis;  
 KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
 KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
 KW diabetic retinopathy; chronic inflammation; cyclic.

XX Synthetic.

XX Key Location/Qualifiers  
 XX Disulfide-bond 1..11  
 FT /note= "This bond cyclises the peptide"

XX WO200152875-A1.  
 PN

XX 26-JUL-2001.  
 PD

XX 18-JAN-2001; 2001WO-US0001533.  
 PF

XX 18-JAN-2000; 2000US-0176293P.  
 PR

XX 16-MAY-2000; 2000US-0204590P.  
 PR

XX

PA (LUDW-) LUDWIG INST CANCER RES.  
 XX Achen MG, Hughes RA, Stacker S, Cendron A;  
 XX WPI; 2001-442248/47.  
 XX Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
 PT or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
 PT from an exposed loop of a growth factor protein by oxidizing the cysteine  
 PT residues.  
 XX  
 XX Claim 49; Page 32; 102pp; English.  
 XX  
 CC The sequence represents a monomeric monocyclic peptide of the invention,  
 CC whose 3-dimensional structure is modelled on the expose loop of human  
 CC VEGFD (vascular endothelial growth factor). The invention relates to a  
 CC method of producing a monomeric monocyclic peptide by a measuring beta-  
 CC beta carbon separation distances on opposite antiparallel strands of a  
 CC peptide loop fragment from an exposed loop of a growth factor protein and  
 CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
 CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
 CC peptides) and a cyclic peptide with at least one amino acid deleted prior  
 CC to cyclisation are used to interfere with angiogenesis,  
 CC neovascularisation or lymphangiogenesis in a mammal with a condition  
 CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
 CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
 CC hemangioma, vascularised malignant or benign tumour, post-recovery  
 CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
 CC trauma, substance-induced neovascularisation of the liver, excessive  
 CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
 CC sequelae, hypertension induced neovascular sequelae, or chronic liver  
 CC infection. The peptides are also used to modulate vascular permeability  
 CC in a mammal (the mammal has a condition characterised by fluid  
 CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
 CC or brain. The peptides are used to image blood vessels and lymphatic  
 CC vasculature. The monomeric and bicyclic peptides are used to interfere  
 CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
 CC are also used in combination with an anti-inflammatory agent, to treat a  
 CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
 CC diabetic retinopathy  
 XX  
 XX Sequence 11 AA;  
 SQ

Query Match 78.8%; Score 41; DB 4; Length 11;  
 Best Local Similarity 81.8%; Pred. No. 2.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 2; Gaps 1;  
 QY 1 CTSVPL--VPC 9  
 DB 1 CTSVPLTSVPC 11  
 ||||| |||  
 ||||| |||

RESULT 4  
 AAR66216  
 ID AAR66216 standard; protein; 342 AA.  
 XX  
 AC AAR66216;  
 XX  
 XX 16-OCT-2003 (revised)  
 DT 04-AUG-1995 (first entry)  
 XX  
 DE Nocardia corallina reductase (encoded by amoB).  
 XX  
 XX alkene monooxygenase; reductase; epoxidation; amod.  
 XX  
 OS Gordonia rubripertinctus.  
 XX  
 PN JP06292571-A.  
 XX  
 XX 21-OCT-1994.  
 PD  
 XX  
 PF 06-APR-1993; 93JP-00105171.  
 XX

PR 06-APR-1993; 93JP-00105171.  
 XX (NIHA ) JAPAN ENERGY CORP.  
 XX  
 DR WPI; 1995-009069/02.  
 XX N-PSDB; AAR79569.  
 XX  
 PT Alkene monooxygenase and corresp. gene - useful for the epoxidation of an  
 PT alkene.  
 XX  
 XX Claim 2; Page 4-5; 30pp; Japanese.  
 XX  
 CC E.coli transformed with the DNA sequence AAR79569 are able to catalyse  
 CC the epoxidation of alkenes. The DNA is derived from Nocardia corallina  
 CC and comprises 4 open reading frames. ORFs amoA and amod encode subunits 1  
 CC and 2 of the alkene monooxygenase enzyme and ORF amod encodes a reductase  
 CC capable of transferring electrons from NADH coenzyme to a monooxygenase.  
 CC (Updated on 16-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 342 AA;  
 Query Match 78.8%; Score 41; DB 2; Length 342;  
 Best Local Similarity 77.8%; Pred. No. 86;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CTSVPLVPC 9  
 DB 75 CASVPLEPC 83  
 ||||| |||  
 ||||| |||

RESULT 5  
 AAR81472  
 ID AAR81472 standard; protein; 342 AA.  
 XX  
 AC AAR81472;  
 XX  
 DT 16-OCT-2003 (revised)  
 DT 07-AUG-1996 (first entry)  
 XX  
 DE Nocardia corallina alkene mono-oxygenase gene product, amod.  
 XX  
 XX Alkene mono-oxygenase; indole; indigo production; biosynthesis;  
 KW microbial oxidation; dye.  
 XX  
 OS Gordonia rubripertinctus; B-276.  
 XX  
 XX JP08023988-A.  
 PN  
 XX 30-JAN-1996.  
 PD  
 XX  
 XX 08-JUL-1994; 94JP-00179688.  
 PF  
 XX 08-JUL-1994; 94JP-00179688.  
 PR  
 XX (NIHA ) JAPAN ENERGY CORP.  
 XX  
 DR WPI; 1996-133426/14.  
 DR N-PSDB; AAT17418.  
 XX  
 XX Prepn. of indigo by a microbiological method - by culturing a microbe  
 PT having alkene monooxygenase activity to oxidise indole to indigo.  
 XX  
 XX Claim 5; Page 8-9; 11pp; Japanese.  
 XX  
 CC AAR81469-R81472 are protein products of the Nocardia corallina strain B-  
 CC 276 alkene mono-oxygenase gene. The gene encodes 4 protein products amoA,  
 CC amoB, amod and amod derived from the 3 different reading frames of the  
 CC operon. The gene is useful for the production of indigo via oxidation of  
 CC indole. Nocardia corallina can be cultured in a medium contg. indole and  
 CC will readily oxidise the indole yielding indigo into the culture medium.  
 CC E. coli may also be transformed with the alkene mono-oxygenase gene and  
 CC used as above to efficiently produce indigo by microbial oxidation.  
 CC (Updated on 16-OCT-2003 to standardise OS field)

XX  
SQ Sequence 342 AA;  
Query Match 78.8%; Score 41; DB 2; Length 342;  
Best Local Similarity 77.8%; Pred. No. 86;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CISVPLVPC 9  
| | | | |  
Db 75 CASVPLEPC 83

RESULT 6  
AAU04545  
ID AAU04545 standard; peptide; 11 AA.  
XX  
AC AAU04545;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE VEGF based monocyclic peptide 23.  
XX  
KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
KW diabetic retinopathy; chronic inflammation; cyclic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Disulfide-bond 1..11 /note= "This bond cyclises the peptide"  
FT  
XX  
PN WO200152875-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 18-JAN-2001; 2001WO-US001533.  
XX  
PR 18-JAN-2000; 2000US-0176293P.  
PR 16-MAY-2000; 2000US-0204590P.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
PI Achen MG, Hughes RA, Stacker S, Cendron A;  
XX  
PS WPI; 2001-442248/47.  
XX  
CC The sequence represents a monomeric monocyclic peptide of the invention,  
CC whose 3-dimensional structure is modelled on the expose loop of human  
CC VEGF (vascular endothelial growth factor). The invention relates to a  
CC method of producing a monomeric monocyclic peptide by a measuring beta-  
CC beta carbon separation distances on opposite antiparallel strands of a  
CC peptide loop fragment from an exposed loop of a growth factor protein and  
CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
CC peptides), and a cyclic peptide with at least one amino acid deleted prior  
CC to cyclisation are used to interfere with angiogenesis,  
CC neovascularisation or lymphangiogenesis in a mammal with a condition  
CC characterised by angiogenesis, neovascularisation or lymphangiogenesis.  
CC The condition is diabetic retinopathy, psoriasis, arthropathy,  
CC hemangioma, vascularised malignant or benign tumour, post-recovery  
CC cerebrovascular accident, post-angioplasty restenosis, head, heat or cold  
CC trauma, substance-induced neovascularisation of the liver, excessive  
CC hormone-related angiogenic dysfunction, diabetes induced neovascular  
CC sequelae, hypertension induced neovascular sequelae, or chronic liver

CC infection. The peptides are also used to modulate vascular permeability  
CC in a mammal (the mammal has a condition characterised by fluid  
CC accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura,  
CC or brain. The peptides are used to image blood vessels and lymphatic  
CC vasculature. The monomeric and bicyclic peptides are used to interfere  
CC with at least one biological activity induced by VEGF, VEGF-C or -D and  
CC are also used in combination with an anti-inflammatory agent, to treat a  
CC chronic inflammation, especially rheumatoid arthritis, psoriasis and  
CC diabetic retinopathy  
XX  
SQ Sequence 11 AA;  
Query Match 76.9%; Score 40; DB 4; Length 11;  
Best Local Similarity 72.7%; Pred. No. 4.2;  
Matches 8; Conservative 1; Mismatches 0; Indels 2; Gaps 1;  
QY 1 CISVPL--VPC 9  
| | | | |  
Db 1 CVSVP LTTVPC 11

RESULT 7  
AAU04542  
ID AAU04542 standard; peptide; 11 AA.  
XX  
AC AAU04542;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE VEGF based monocyclic peptide 20.  
XX  
KW Human; VEGF; vascular endothelial growth factor; angiogenesis;  
KW neovascularisation; lymphangiogenesis; psoriasis; tumour;  
KW diabetes induced neovascular sequelae; rheumatoid arthritis;  
KW diabetic retinopathy; chronic inflammation; cyclic.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Disulfide-bond 1..11 /note= "This bond cyclises the peptide"  
FT  
XX  
PN WO200152875-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 18-JAN-2001; 2001WO-US001533.  
XX  
PR 18-JAN-2000; 2000US-0176293P.  
PR 16-MAY-2000; 2000US-0204590P.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
PI Achen MG, Hughes RA, Stacker S, Cendron A;  
XX  
PS WPI; 2001-442248/47.  
XX  
CC Novel monomeric monocyclic peptide, used to interfere with angiogenesis,  
CC or lymphangiogenesis, is produced by cyclizing a peptide loop fragment  
CC from an exposed loop of a growth factor protein by oxidizing the cysteine  
CC residues.  
XX  
PS Example 25; Page 47; 102pp; English.  
XX  
CC The sequence represents a monomeric monocyclic peptide of the invention,  
CC whose 3-dimensional structure is modelled on the expose loop of human  
CC VEGF (vascular endothelial growth factor). The invention relates to a  
CC method of producing a monomeric monocyclic peptide by a measuring beta-  
CC beta carbon separation distances on opposite antiparallel strands of a  
CC peptide loop fragment from an exposed loop of a growth factor protein and  
CC cyclising the peptide by oxidising the cysteine residues. The monocyclic  
CC peptides, dimeric bicyclic peptides (comprising 2 linked monocyclic  
CC peptides), and a cyclic peptide with at least one amino acid deleted prior

to cyclisation are used to interfere with angiogenesis, neovascularisation or lymphangiogenesis in a mammal with a condition characterised by angiogenesis, neovascularisation or lymphangiogenesis. The condition is diabetic retinopathy, psoriasis, arthropathy, hemangioma, vascularised malignant or benign tumour, post-recovery cerebrovascular accident, post-angioplasty restenosis, head, heat or cold trauma, substance-induced neovascularisation of the liver, excessive hormone-related angiogenic dysfunction, diabetes induced neovascular sequelae, hypertension induced neovascular sequelae, or chronic liver infection. The peptides are also used to modulate vascular permeability in a mammal (the mammal has a condition characterised by fluid accumulation in peripheral limbs or in lungs, peritoneal cavity, pleura, or brain). The peptides are used to image blood vessels and lymphatic vasculature. The monomeric and bicyclic peptides are used to interfere with at least one biological activity induced by VEGF, VEGF-C or -D and are also used in combination with an anti-inflammatory agent, to treat a chronic inflammation, especially rheumatoid arthritis, psoriasis and diabetic retinopathy.

Sequence 11 AA;

Query Match 75.0%; Score 39; DB 4; Length 11;  
Best Local Similarity 72.7%; Pred. No. 6.2;  
Matches 8; Conservative 1; Mismatches 0; Indels

QY	1	CISVPL--VPC	9
		:	
Db	1	CLSVPLTSVPC	11

## RESULT 8

AAM06501  
 ID AAM06501 standard; protein; 101 AA.

AA  
AC

AA	
DT	05-OCT-2001 (first entry)

Human foetal protein, SEO ID NO: 232.

Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
 neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
 gene therapy; antisense therapy; cancer; immune disorder;  
 growth disorder; osteoporosis; thrombolytic disorder;  
 nervous system disorder; inflammation.

OS Homo sapiens.

AA  
PN  
WO200155339-A2.

XX  
PD 02-AUG-2001.

XX  
PF 25--TAN-2001 : 2001MO-VIS0002723

XX  
PP 25-JAN-2000. 2000UTC 004031404

PR 15-SEP-2000; 2000US-00663870.  
PR 05 NOV 2000; 2000US-00663870.  
PR 05 NOV 2000; 2000US-00663870.

XX  
XX

[illegible]

PI Ieung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;  
PI Liu C, Asundi V, Zhou P, Werhman T;

XX  
DR  
WPI: 2001-465571/50.

DR N-PSDB; AAH94176.  
XX

PT Novel fetal proteins useful for the treatment and diagnosis of diseases  
PT associated with dysfunction of the protein e.g. cancers, immune  
PT disorders, growth disorders, thrombolytic disorders, nervous system  
PT disorders and inflammation.

PS Claim 10; Page 264; 715pp; English.

XX

CC The invention relates to novel foetal polypeptides encoded by  
CC polynucleotides comprising one of 477 sequences fully defined in the  
CC specification. The foetal polynucleotides and polypeptides are useful in  
CC the treatment and diagnosis of diseases such as cancers, immune  
CC disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,  
CC nervous system disorders and inflammation. The present sequence is a  
CC polypeptide encoded by a cDNA assembled using an expressed sequence tag  
CC (EST) found to be expressed in human foetal tissue cDNA libraries

Sequence 101 AA;

Query Match 75.0%; Score 39; DB 4; Length 101;  
Best Local Similarity 87.5%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 1; Indels

Qy	1	2	8	9
Db	CISVPLVP	CISVPLTP		

RESULT 9  
ABG22400

ID ABG22400 standard; protein; 102 AA.

AC ABG22400;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #22391.

XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX  
PN  
WO200175067-A2.

XX 11-OCT-2007 PD

XX  
PF 30-MAR-2001; 2001WO-US008631.

XX  
PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.  
XX

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

AA  
DR WPI; 2001-639362/73.

DR N-PSDB; AAS86587.

xx New isolated polynucleotide and encoded polypeptides, useful in  
pt diagnostics, forensics, gene mapping, identification of mutations  
pt responsible for genetic disorders or other traits and to assess  
pt biodiversity.

PS Claim 20; SEQ ID NO 52759; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity.



DE Chlamydia psittaci infection-specific protein IncB.  
 XX IncB; infection; vaccine; therapy; diagnosis.  
 XX Chlamydomydia caviae.  
 XX WO9953948-A1.  
 XX 28-OCT-1999.  
 XX 20-APR-1999; 99WO-US008744.  
 XX 20-APR-1998; 98US-0082438P.  
 PR 21-APR-1998; 98US-0082588P.  
 PR 22-MAY-1998; 98US-0086450P.  
 XX (UYOR-) UNIV OREGON STATE.  
 XX Rocky DD, Bannantine JP;  
 XX WPI; 1999-633904/54.  
 DR N-PSDB; AAZ34588.  
 XX Novel bacterial infection specific proteins for treating and diagnosing  
 PT chlamydial infections.  
 XX Claim 1; Page 41-42; 56pp; English.  
 XX This sequence represents novel infection-specific protein IncB of  
 CC Chlamydia psittaci strain GPIC. IncB is found in the inclusion membrane  
 CC of infected cells. It is associated primarily with the vegetative  
 CC reticulate body form of Chlamydia rather than with the refractile  
 CC elementary body form. The invention includes: a vaccine directed against  
 CC the reticulate body form of Chlamydia comprising 1 or more infection-  
 CC specific proteins (see AAY32170-78), including Inca, IncB and IncC;  
 CC methods of using and producing such a vaccine; methods for detection of  
 CC infection-specific antibodies or antigens in a biological specimen; and a  
 CC method of using therapeutic agents specifically directed against  
 CC infection-specific peptides, or the genes that code for such peptides, to  
 CC treat chlamydial infection. (Updated on 17-OCT-2003 to standardise OS  
 CC field)  
 XX Sequence 198 AA;  
 SQ Query Match 73.1%; Score 38; DB 2; Length 198;  
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CUSVPLVPC 9  
 DB 157 CUSLPIISC 165  
 RESULT 13  
 ADB64366  
 ID ADB64366 standard; protein; 208 AA.  
 XX ADB64366;  
 AC ADB64366;  
 XX 04-DEC-2003 (first entry)  
 DE Human protein encoded by clone FCBBF30171230.  
 XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumour.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX EP1308459-A2.  
 PN EP1308459-A2.  
 XX 07-MAY-2003.  
 PD 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.  
 PF 05-NOV-2001; 2001JP-00379298.  
 XX 25-JAN-2002; 2002US-00350978.  
 PR (HELI-) HELIX RES INST.  
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
 PA Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 PI WPI; 2003-450961/43.  
 DR N-PSDB; ADB62396.  
 XX New polynucleotides and polypeptides, useful for developing a diagnostic  
 PT marker or medicines for regulation of their expression and activity, or  
 PT as targets of gene therapy.  
 XX Claim 1; Page; 222pp; English.  
 XX The invention discloses a polynucleotide comprising a sequence selected  
 CC from 1970 fully defined nucleotide sequences which encode novel  
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 CC or its partial peptide, an antibody binding to the polypeptide or peptide  
 CC of the polynucleotide, immunologically assaying the polypeptide or  
 CC peptide of the polynucleotide by contacting the polypeptide or peptide  
 CC with the antibody of the encoded protein, and observing the binding  
 CC between the two, a transformant carrying the polynucleotide in an  
 CC expressible manner and an antisense polynucleotide. The oligonucleotide  
 CC is useful as a primer for synthesising the polynucleotide, or as a probe  
 CC for detecting the polynucleotide. The polynucleotides and encoded  
 CC proteins are useful as pharmaceutical agents and many disease-related  
 CC genes may be included in them, for developing a diagnostic marker or  
 CC medicines for regulation of their expression and activity, or as targets  
 CC of gene therapy. The genes are involved in tissue and/or cell  
 CC regeneration. Membrane proteins, signal transduction-related proteins,  
 CC transcription-related proteins, disease-related proteins and genes  
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate  
 CC the activity or expression of the encoded protein to treat diseases. The  
 CC sequence presented is a protein of the invention. Note: Some of the  
 CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX Sequence 208 AA;  
 SQ Query Match 73.1%; Score 38; DB 7; Length 208;  
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CUSVPLVPC 9  
 DB 33 CASLPLSPC 41  
 RESULT 14  
 ABB57779  
 ID ABB57779 standard; protein; 562 AA.  
 XX ABB57779;  
 AC ABB57779;  
 XX 26-MAR-2002 (first entry)  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster polypeptide SEQ ID NO 129.  
 DE Drosophila melanogaster polypeptide SEQ ID NO 129.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW Drosophila melanogaster.  
 OS Drosophila melanogaster.  
 XX Drosophila melanogaster.

```

PN WO200171042-A2.
XX
XX
XX 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL01882.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX PS Disclosure; SEQ ID NO 129; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 562 AA;

Query Match 73.1%; Score 38; DB 4; Length 562;
Best Local Similarity 55.6%; Pred. No. 4.4e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLVPC 9
|:|:|
Db 9 CVKVPIDPC 17

RESULT 15
ADA35614
ID ADA35614 standard; protein; 589 AA.
XX
XX AC ADA35614;
XX
XX DT 20-NOV-2003 (first entry)
XX
XX DE Acinetobacter baumannii protein #2775.
XX
XX KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX plant biocontrol agent.
XX
XX OS Acinetobacter baumannii.
XX
XX PN US6562958-B1.
XX
XX PD 13-MAY-2003.
XX
XX PF 04-JUN-1999; 99US-00328352.
XX
XX PR 09-JUN-1998; 98US-0088701P.
XX
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PI Breton G, Bush D;
XX
XX WPI; 2003-576092/54.
XX DR N-PSDB; ADA31488.

```

```

XX
XX PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX plants.
XX
XX PS Example; SEQ ID NO 6901; 328pp; English.
XX
XX CC The invention relates to isolated Acinetobacter baumannii nucleic acids.
XX The A. baumannii nucleic acids and polypeptides are useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, to detect the presence of
XX A. baumannii and other Acinetobacter species in a sample, in screening
XX compounds for the ability to interfere with the A. baumannii life cycle
XX or to inhibit A. baumannii infection, and as biocontrol agents for
XX plants. The present sequence represents the amino acid sequence of an A.
XX baumannii protein.
XX
XX SQ Sequence 589 AA;

Query Match 73.1%; Score 38; DB 6; Length 589;
Best Local Similarity 77.8%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLVPC 9
|:|:|
Db 273 C1SVPLPHC 281

Search completed: September 5, 2004, 09:55:23
Job time : 31.5455 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 10:00:15 ; Search time 25,2727 Seconds  
(without alignments)  
112.199 Million cell updates/sec

Title: US-09-761-636A-14  
Perfect score: 52  
Sequence: 1 CISVPLVPC 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1298764 seqs, 315065143 residues

Total number of hits satisfying chosen parameters: 1298764

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

#### Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	9	US-09-761-636A-14	Sequence 14, Appl
2	41.5	79.8	10	US-09-761-636A-13	Sequence 13, Appl
3	41	78.8	11	US-09-761-636A-7	Sequence 7, Appl
4	40	76.9	11	US-09-761-636A-26	Sequence 26, Appl
5	39	75.0	11	US-09-761-636A-23	Sequence 23, Appl
6	39	75.0	71	US-10-437-963-181998	Sequence 181998,
7	38	73.1	208	US-10-104-047-2520	Sequence 2520, Ap
8	38	73.1	413	US-10-087-192-519	Sequence 519, App
9	37	71.2	87	US-10-424-599-258586	Sequence 258586,
10	37	71.2	94	US-10-424-599-165772	Sequence 165772,
11	37	71.2	149	US-10-424-599-158427	Sequence 158427,
12	37	71.2	310	US-10-094-749-3057	Sequence 3057, Ap
13	36	69.2	11	US-09-761-636A-25	Sequence 25, Appl
14	36	69.2	53	US-10-437-963-148570	Sequence 148570,
15	36	69.2	60	US-10-424-599-279501	Sequence 279501,

16	36	69.2	216	16	US-10-437-963-159284	Sequence 159284,
17	36	69.2	689	16	US-10-437-963-176796	Sequence 176796,
18	36	69.2	725	16	US-10-437-963-187630	Sequence 187630,
19	36	69.2	815	15	US-10-369-493-22558	Sequence 22558, A
20	35.5	68.3	89	12	US-10-424-599-255153	Sequence 255153,
21	35	67.3	56	13	US-10-114-893-131	Sequence 131, App
22	35	67.3	58	12	US-10-424-599-225406	Sequence 225406,
23	35	67.3	62	12	US-10-424-599-162402	Sequence 162402,
24	35	67.3	92	16	US-10-437-963-162997	Sequence 162997,
25	35	67.3	93	12	US-10-424-599-267602	Sequence 267602,
26	35	67.3	233	12	US-10-425-114-43943	Sequence 43943, A
27	35	67.3	233	12	US-10-425-114-44103	Sequence 44103, A
28	35	67.3	242	12	US-10-425-114-43762	Sequence 43762, A
29	35	67.3	249	9	US-09-745-763-9	Sequence 9, Appli
30	35	67.3	249	12	US-10-147-493-110	Sequence 110, App
31	35	67.3	249	12	US-10-145-127-110	Sequence 110, App
32	35	67.3	249	12	US-10-160-503-110	Sequence 110, App
33	35	67.3	249	12	US-10-143-118-110	Sequence 110, App
34	35	67.3	249	12	US-10-144-993-110	Sequence 110, App
35	35	67.3	249	12	US-10-158-787-110	Sequence 110, App
36	35	67.3	249	12	US-10-140-024-110	Sequence 110, App
37	35	67.3	249	12	US-10-140-808-110	Sequence 110, App
38	35	67.3	249	12	US-10-152-405-110	Sequence 110, App
39	35	67.3	249	12	US-10-127-852A-110	Sequence 110, App
40	35	67.3	249	12	US-10-127-900A-110	Sequence 110, App
41	35	67.3	249	12	US-10-128-685A-110	Sequence 110, App
42	35	67.3	249	12	US-10-131-820A-110	Sequence 110, App
43	35	67.3	249	12	US-10-142-886-110	Sequence 110, App
44	35	67.3	249	12	US-10-146-728-110	Sequence 110, App
45	35	67.3	249	12	US-10-146-786-110	Sequence 110, App

#### ALIGNMENTS

#### RESULT 1

US-09-761-636A-14  
; Sequence 14, Application US/09761636A  
; Patent No. US20020065218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09/761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-761-636A-14

Query Match 100.0%; Score 52; DB 9; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9  
Db 1 CISVPLVPC 9

#### RESULT 2

US-09-761-636A-13  
; Sequence 13, Application US/09761636A  
; Patent No. US20020065218A1

```

; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-761-636A-13

Query Match          79.8%; Score 41.5; DB 9; Length 10;
Best Local Similarity 90.0%; Pred. No. 2.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 1;

Qy 1 C1SVPL-VPC 9
Db 1 C1SVPLSVPC 10

RESULT 3
US-09-761-636A-7
; Sequence 7, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-761-636A-7

Query Match          78.8%; Score 41; DB 9; Length 11;
Best Local Similarity 81.8%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 1;

Qy 1 C1SVPL-VPC 9
Db 1 C1SVPLSVPC 11

RESULT 4
US-09-761-636A-26
; Sequence 26, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 11
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-761-636A-23

Query Match          75.0%; Score 39; DB 9; Length 11;
Best Local Similarity 72.7%; Pred. No. 7.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 1;

Qy 1 C1SVPL-VPC 9
Db 1 C1SVPLSVPC 11

RESULT 5
US-09-761-636A-23
; Sequence 23, Application US/09761636A
; Patent No. US20020065218A1
; GENERAL INFORMATION:
; APPLICANT: ACHEN, Marc
; APPLICANT: STACKER, Steven
; APPLICANT: HUGHES, Richard
; APPLICANT: CENDRON, Angela
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR
; FILE REFERENCE: 1064/48505 Achen et al
; CURRENT APPLICATION NUMBER: US/09/761,636A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/176,293
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/204,590
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 11
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-761-636A-23

Query Match          76.9%; Score 40; DB 9; Length 11;
Best Local Similarity 72.7%; Pred. No. 5;
Matches 8; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

Qy 1 C1SVPL-VPC 9
Db 1 C1SVPLSVPC 11

RESULT 6
US-10-437-963-181998
; Sequence 181998, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 181998  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_79226C.1.pep  
US-10-437-963-181998

Query Match 75.0%; Score 39; DB 16; Length 71;  
Best Local Similarity 62.5%; Pred. No. 42;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CJSVPLVP 8  
|:|:|:|  
Db 57 CLSIPLIP 64

## RESULT 7

US-10-104-047-2520  
; Sequence 2520, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1el full length cdna  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR FILING DATE:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2520  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2520

Query Match 73.1%; Score 38; DB 15; Length 208;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CJSVPLVPC 9  
|:|:|:|  
Db 33 CASLPLSPC 41

## RESULT 8

US-10-087-192-519  
; Sequence 519, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 519  
; LENGTH: 413  
; TYPE: PRT  
; ORGANISM: Mus musculus

; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(413)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-087-192-519

Query Match 73.1%; Score 38; DB 12; Length 413;  
Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CJSVPLVPC 9  
|:|:|:|  
Db 287 CSQAPLVPC 295

## RESULT 9

US-10-424-599-258586  
; Sequence 258586, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 258586  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_75528C.1.pep  
US-10-424-599-258586

Query Match 71.2%; Score 37; DB 12; Length 87;  
Best Local Similarity 55.6%; Pred. No. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CJSVPLVPC 9  
|:|:|:|  
Db 69 CLSIQLTPC 77

## RESULT 10

US-10-424-599-165772  
; Sequence 165772, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 165772  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(94)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_120707C.1.pep  
US-10-424-599-165772

Query Match 71.2%; Score 37; DB 12; Length 94;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9  
|:|:|:|:|  
Db 33 CVSRPGVPC 41

RESULT 11  
US-10-424-599-158427  
; Sequence 158427, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 158427  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(149)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_114078C.1.pep  
US-10-424-599-158427

Query Match 71.2%; Score 37; DB 12; Length 149;  
Best Local Similarity 66.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9  
|:|:|:|:|  
Db 21 CRSCPVPC 29

RESULT 12  
US-10-094-749-3057  
; Sequence 3057, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOTYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3057  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-094-749-3057

Query Match 71.2%; Score 37; DB 15; Length 310;  
Best Local Similarity 55.6%; Pred. No. 3.5e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CISVPLVPC 9  
|:|:|:|:|  
Db 202 CLSLPQAPC 210

RESULT 13  
US-09-761-636A-25  
; Sequence 25, Application US/09761636A  
; Patent No. US20020065218A1  
; GENERAL INFORMATION:  
; APPLICANT: ACHEN, Marc  
; APPLICANT: STACKER, Steven  
; APPLICANT: HUGHES, Richard  
; APPLICANT: CENDRON, Angela  
; TITLE OF INVENTION: VEGF-D/VEGF-C/VEGF PEPTIDOMIMETIC INHIBITOR  
; FILE REFERENCE: 1064/48505 Achen et al  
; CURRENT APPLICATION NUMBER: US/09/761,636A  
; CURRENT FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 60/176,293  
; PRIOR FILING DATE: 2000-01-18  
; PRIOR APPLICATION NUMBER: US 60/204,590  
; PRIOR FILING DATE: 2000-05-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 25  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: synthetic construct  
US-09-761-636A-25

Query Match 69.2%; Score 36; DB 9; Length 11;  
Best Local Similarity 63.6%; Pred. No. 22;  
Matches 7; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

Qy 1 CISVPL-VPC 9  
|:|:|:|:|  
Db 1 CISLPISVPC 11

RESULT 14  
US-10-437-963-148570  
; Sequence 148570, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14

```
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 148570
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_48989C.1.pep
US-10-437-963-148570

Query Match          69.2%; Score 36; DB 16; Length 53;
Best Local Similarity 55.6%; Pred. No. 98;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      1 C1SVPLVPC 9
Db      13 CLSVLVIPC 21

RESULT 15
US-10-424-599-279501
; Sequence 279501, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 279501
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_94411C.1.pep
US-10-424-599-279501

Query Match          69.2%; Score 36; DB 12; Length 60;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 C1SVPLVLP 8
Db      28 CMKVPLVLP 35
```

Search completed: September 5, 2004, 10:29:24  
Job time : 26.2727 secs

***This Page Blank (uspto)***

***Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2004, 09:55:30 ; Search time 8 Seconds  
(without alignments)  
58.079 Million cell updates/sec

Title: US-09-761-636A-14

Perfect score: 52

Sequence: 1 CISVPLVPC 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCFUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	78.8	342	1	US-08-499-215-5
2	38	73.1	44	4	US-09-341-461-10
3	38	73.1	589	4	US-09-328-352-6901
4	38	73.1	3623	4	US-09-341-461-2
5	37	71.2	168	4	US-09-252-991A-27872
6	36	69.2	455	4	US-09-489-039A-13375
7	35	67.3	90	4	US-09-252-991A-20360
8	35	67.3	1440	3	US-09-357-251-37
9	34	67.3	1512	4	US-09-443-184-48
10	34	65.4	96	4	US-09-540-236-2315
11	34	65.4	147	4	US-09-252-991A-23997
12	34	65.4	215	1	US-08-107-684B-11
13	34	65.4	1876	2	US-08-609-049A-12
14	34	65.4	1876	2	US-08-609-049A-28
15	34	65.4	1876	3	US-09-170-996-12
16	34	65.4	1876	3	US-09-170-996-28
17	33	63.5	20	3	US-09-461-697-169
18	33	63.5	32	3	US-09-461-697-167
19	33	63.5	322	4	US-09-543-681A-4992
20	33	63.5	994	4	US-09-543-681A-7288
21	33	63.5	1253	1	US-07-920-281C-3
22	33	63.5	1253	3	US-08-466-277-3
23	32.5	62.5	62	4	US-09-621-976-4988
24	32	61.5	109	4	US-09-087-031E-19
25	32	61.5	110	4	US-09-621-976-4028
26	32	61.5	111	4	US-09-252-991A-26531
27	32	61.5	180	4	US-09-544-716-19

28 32 61.5 180 4 US-09-557-921-20 Sequence 20, Appl  
29 32 61.5 180 4 US-09-619-380-21 Sequence 21, Appl  
30 32 61.5 184 4 US-09-489-039A-11648 Sequence 11648, A  
31 32 61.5 185 1 US-07-988-273-2 Sequence 2, Appl  
32 61.5 185 3 US-08-848-810-25 Sequence 25, Appl  
33 32 61.5 185 3 US-09-164-193-21 Sequence 21, Appl  
34 32 61.5 185 4 US-09-221-448A-21 Sequence 21, Appl  
35 32 61.5 185 5 PCT-US93-12019-2 Sequence 2, Appl  
36 32 61.5 195 4 US-09-546-043-5 Sequence 5, Appl  
37 32 61.5 218 3 US-08-985-526-1 Sequence 1, Appl  
38 32 61.5 239 5 PCT-US93-01652-1 Sequence 1, Appl  
39 32 61.5 246 4 US-09-546-043-6 Sequence 6, Appl  
40 32 61.5 267 4 US-09-546-043-7 Sequence 7, Appl  
41 32 61.5 305 3 US-09-248-588-13 Sequence 13, Appl  
42 32 61.5 313 4 US-09-087-031E-4 Sequence 4, Appl  
43 32 61.5 313 4 US-09-546-043-3 Sequence 3, Appl  
44 32 61.5 313 4 US-09-514-885-1 Sequence 1, Appl  
45 32 61.5 314 4 US-08-937-067-7 Sequence 7, Appl

## ALIGNMENTS

RESULT 1  
US-08-499-215-5  
; Sequence 5, Application US/08499215  
; Patent No. 5612204  
; GENERAL INFORMATION:  
; APPLICANT: Saeki, Hiasashi  
; APPLICANT: Miura, Akira  
; TITLE OF INVENTION: BIOLOGICAL DEGRADATIVE  
; TREATMENT OF CHLORINE-SUBSTITUTED ETHYLENE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W.  
; STATE: Washington D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; SOFTWARE: & WordPerfect version 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/499,215  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP Hei-6-179689  
; FILING DATE: 08-JUL-1994  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 342 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-499-215-5

Query Match 78.8%; Score 41; DB 1; Length 342;  
Best Local Similarity 77.8%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CISVPLVPC 9

Db 75 CASVPLEPC 83

RESULT 2

US-09-341-461-10  
; Sequence 10, Application US/09341461

```
; Patent No. 6586389
; GENERAL INFORMATION:
; APPLICANT: Hammond, Timothy G.
; APPLICANT: Verroust, Pierre J.
; TITLE OF INVENTION: Cubilin Protein, DNA Sequences Encoding Cubilin
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: D6148
; CURRENT APPLICATION NUMBER: US/09/341,461
; CURRENT FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: PCT/US99/01259
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 40
; SEQ ID NO 10
; LENGTH: 44
; TYPE: PRT
; ORGANISM: rat
; FEATURE:
; OTHER INFORMATION: amino acid sequence of rat cubilin EGF5 repeat
US-09-341-461-10

Query Match      73.1%; Score 38; DB 4; Length 44;
Best Local Similarity 66.7%; Pred. No. 9.2;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CISVPLVPC 9
Db      12 CSQAPLVPC 20

RESULT 3
US-09-328-352-6901
; Sequence 6901, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6901
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6901

Query Match      73.1%; Score 38; DB 4; Length 589;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CISVPLVPC 9
Db      273 CISVPLFHC 281

RESULT 4
US-09-341-461-2
; Sequence 2, Application US/09341461
; Patent No. 6586389
; GENERAL INFORMATION:
; APPLICANT: Hammond, Timothy G.
; APPLICANT: Verroust, Pierre J.
; TITLE OF INVENTION: Cubilin Protein, DNA Sequences Encoding Cubilin
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: D6148
; CURRENT APPLICATION NUMBER: US/09/341,461
; CURRENT FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: PCT/US99/01259
; PRIOR FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 40
; SEQ ID NO 2
; LENGTH: 3623
; TYPE: PRT
; ORGANISM: rat
; FEATURE:
; OTHER INFORMATION: amino acid sequence of rat cubilin protein
US-09-341-461-2

Query Match      73.1%; Score 38; DB 4; Length 3623;
Best Local Similarity 66.7%; Pred. No. 7.7e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CISVPLVPC 9
Db      313 CSQAPLVPC 321

RESULT 5
US-09-252-991A-27872
; Sequence 27872, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27872
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27872

Query Match      71.2%; Score 37; DB 4; Length 168;
Best Local Similarity 55.6%; Pred. No. 51;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CISVPLVPC 9
Db      37 CTSIPGIPC 45

RESULT 6
US-09-489-039A-13375
; Sequence 13375, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13375
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13375

Query Match      69.2%; Score 36; DB 4; Length 455;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CISVPLVPC 9
Db      8 CYFFLVPC 16
```



```

RESULT 7
US-09-252-991A-20960
; Sequence 20960, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20960
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20960
Query Match 67.3%; Score 35; DB 4; Length 90;
Best Local Similarity 55.6%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 C1SVPLVPC 9
Db 12 CASCPTPC 20

RESULT 8
US-09-357-251-37
; Sequence 37, Application US/09357251
; Patent No. 6271441
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Lavo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Schwaber, James S.
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
; FILE REFERENCE: BB-1193
; CURRENT APPLICATION NUMBER: US/09/357,251
; CURRENT FILING DATE: 1999-07-20
; EARLIER APPLICATION NUMBER: 60/093,530
; EARLIER FILING DATE: July 21, 1998
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
; LENGTH: 1440
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-357-251-37
Query Match 67.3%; Score 35; DB 3; Length 1440;
Best Local Similarity 44.4%; Pred. No. 9.3e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 C1SVPLVPC 9
Db 1229 CVQVWIIPC 1237

RESULT 9
US-09-443-184-48
; Sequence 48, Application US/09443184A
; Patent No. 6372431
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary
; APPLICANT: Kaser, Matthew R.

```

```

; APPLICANT: Panzer, Scott
; APPLICANT: Seilhammer, Jeffrey J.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah
; APPLICANT: Azimzai, Valda
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
; FILE REFERENCE: PC-0007 US
; CURRENT APPLICATION NUMBER: US/09/443,184A
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 48
; LENGTH: 1512
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6372431 2302721CD1
US-09-443-184-48
Query Match 67.3%; Score 35; DB 4; Length 1512;
Best Local Similarity 44.4%; Pred. No. 9.8e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 C1SVPLVPC 9
Db 1301 CVQVWIIPC 1309

RESULT 10
US-09-540-236-2315
; Sequence 2315, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2315
; LENGTH: 96
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2315
Query Match 65.4%; Score 34; DB 4; Length 96;
Best Local Similarity 87.5%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 C1SVPLVP 8
Db 87 C1SVRLVP 94

RESULT 11
US-09-252-991A-23997
; Sequence 23997, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```
; SEQ ID NO 23997
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23997

Query Match          65.4%; Score 34; DB 4; Length 147;
Best Local Similarity 53.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

Qy 1 C1SV-PLVPC 9
Db 105 C1SV-PLVPC 117

RESULT 12
US-08-107-684B-11
; Sequence 11, Application US/08107684B
; Patent No. 5552273
; GENERAL INFORMATION:
; APPLICANT: CLEUZAT, Philippe L.
; APPLICANT: AWADE, Abalo
; APPLICANT: ROBERT-BAUDOUY, Jeannine
; APPLICANT: GAYRAL, Jean-Pierre
; TITLE OF INVENTION: POLYPEPTIDES CONTAINING SEQUENCES
; TITLE OF INVENTION: CHARACTERISTIC OF PYRROLIDONE CARBOXYLYL PEPTIDASES,
; TITLE OF INVENTION: POLYNUCLEOTIDES CONTAINING A SEQUENCE CODING FOR SUCH
; TITLE OF INVENTION: POLYPEPTIDES, AND THEIR USE, IN PARTICULAR FOR DIAGNOSTIC
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESS: OLIFF & BERRIDGE
; STREET: P.O. Box 19928
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,684B
; FILING DATE: 17-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 28478
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6400
; TELEFAX: (703) 836-2787
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-107-684B-11

Query Match          65.4%; Score 34; DB 1; Length 215;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLVP 8
Db 37 C1SVPLVP 44

RESULT 13
US-08-609-049A-12
; Sequence 12, Application US/08609049A
; Patent No. 5948664
```

```
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: NO. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-609-049A-12

Query Match          65.4%; Score 34; DB 2; Length 1876;
Best Local Similarity 55.6%; Pred. No. 1.8e+03;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLVPC 9
Db 848 C1SVPLVPC 856

RESULT 14
US-08-609-049A-28
; Sequence 28, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: NO. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
```

ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 2307K-063700US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1876 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-609-049A-28

Query Match 65.4%; Score 34; DB 2; Length 1876;  
Best Local Similarity 55.6%; Pred. No. 1.8e+03;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLVPC 9  
Db 848 CVSKEPIVVC 856

RESULT 15  
US-09-170-996-12  
; Sequence 12, Application US/09170996  
; Patent No. 6291220  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Lewis T.  
; APPLICANT: Molz, Lisa  
; APPLICANT: Chen, Yen-Wen  
; TITLE OF INVENTION: No. 6291220e1 PI 3-Kinase Polypeptides  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/170,996  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/609,049  
; FILING DATE: 29-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Karen B.  
; REGISTRATION NUMBER: 29,684  
; REFERENCE/DOCKET NUMBER: 2307K-063700US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1876 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-170-996-12

Query Match 65.4%; Score 34; DB 3; Length 1876;  
Best Local Similarity 55.6%; Pred. No. 1.8e+03;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 C1SVPLVPC 9  
Db 848 CVSKEPIVVC 856  
Search completed: September 5, 2004, 10:22:01  
Job time : 8 secs

**THIS PAGE BLANK (USPTO)**